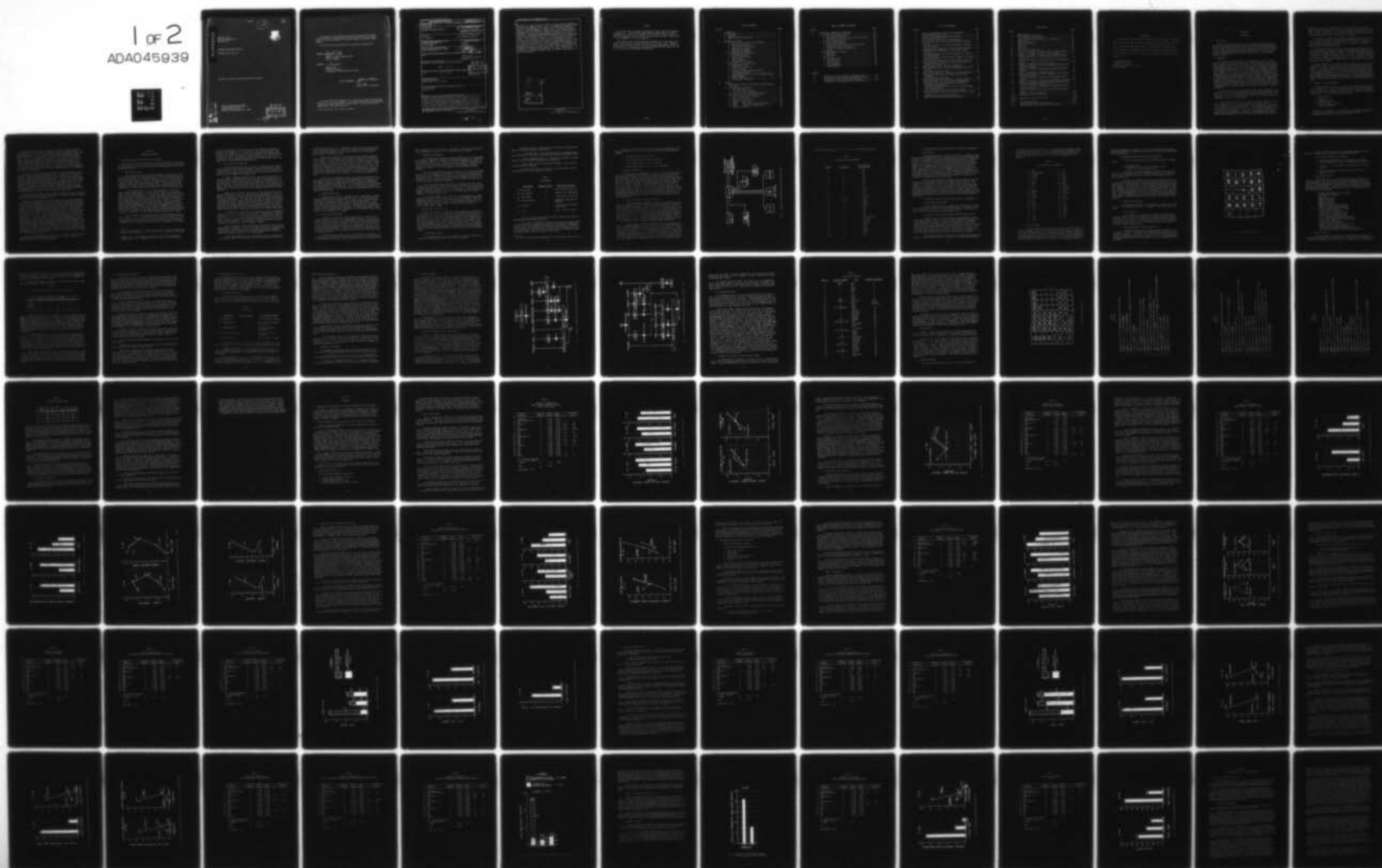
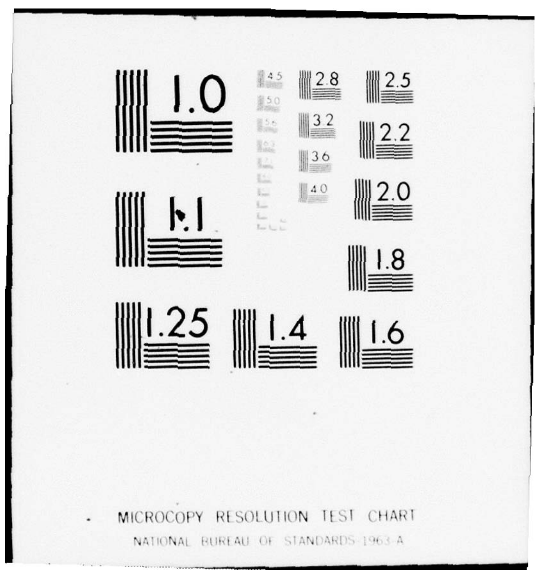
AD-A045 939 THRESHOLD TECHNOLOGY INC DELRAN N J F/G 17/2
AUTOMATIC DATA ENTRY ANALYSIS.(U)
SEP 77 J R WELCH F30602-76-C-0178
UNCLASSIFIED RADC-TR-77-306 NL

1 OF 2
ADA045939




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

12

RADC-TR-77-306
Final Technical Report
September 1977

AUTOMATIC DATA ENTRY ANALYSIS

Threshold Technology Inc.

Approved for public release; distribution unlimited.

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

This report has been reviewed and is approved for publication.

APPROVED: Robert A. Curtis

ROBERT A. CURTIS, Captain, USAF
Project Engineer

APPROVED: H Davis

HOWARD DAVIS
Technical Director
Intelligence & Reconnaissance Division

FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (IRAD) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return this copy. Retain or destroy.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

18

1. REPORT NUMBER: RADC-TR-77-306 19

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

6

4. TITLE (and Subtitle): AUTOMATIC DATA ENTRY ANALYSIS.

9

5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report, 29 Mar 76 — 22 Aug 77,

6. PERFORMING ORG. REPORT NUMBER: N/A

10

7. AUTHOR(s): John R. Welch

15

8. CONTRACT OR GRANT NUMBER(s): F30602-76-C-0178 new

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Threshold Technology, Inc.
1829 Underwood Blvd
Delran NJ 08075

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
62702F
40270801

11. CONTROLLING OFFICE NAME AND ADDRESS:
Rome Air Development Center (IRAD)
Griffiss AFB NY 13441

11

12. REPORT DATE: Sep~~tember~~ 1977

13. NUMBER OF PAGES: ~~121~~ 12 141p.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION DOWNGRADING SCHEDULE: N/A

16. DISTRIBUTION STATEMENT (of this Report):
Approved for public release; distribution unlimited.

16 4027

DDC
RECEIVED
NOV 3 1977
B

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
Same

17 Ø8

18. SUPPLEMENTARY NOTES:
RADC Project Engineer:
Captain Robert A. Curtis (IRAD)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):
Voice Data Entry
Human Factors

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

Speed and accuracy have been compared for isolated-word voice recognition, keyboard, and graphical menu data entry systems. One entry task involved simple copying of numeric and alphanumeric data strings. A second was a simulation of complex flight data entry **scenario.** The factors evaluated included voice response feedback and prompting, hand occupation during data entry, and subject experience. Keyboard provided the fastest and most accurate entry of numeric data strings and the fastest entry of alphanumeric strings by subjects with keyboard experience in the simple scenario, but was over

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

390 798 B

slow relative to voice and graphical menu for entry of words by inexperienced subjects in the complex scenario. Voice entry provided the lowest error rate for entry of alphanumeric data strings in the simple scenario primarily because of its greater immunity to reading errors. In the complex scenario, voice was faster than keyboard for inexperienced subjects, and had a similar operational error rate, but had a substantially higher error rate before correction. Graphical menu ranked between keyboard and voice in most of the simple scenario measures, except that it was least accurate with alphanumeric data, and had the lowest entry speed for long strings. In the complex scenario, the performance of graphical menu differed significantly from that of voice primarily by virtue of its lower error rate before correction. Overall, most of the errors with voice input involved misrecognition while those with keyboard and graphical menu involved misreading. Voice response feedback was too slow to be of value in the simple scenario, but voice response prompting significantly reduced reading errors in the complex scenario. Hand occupation, of substantial duration, gave voice input a relative speed advantage over keyboard and graphical menu, but increased entry time and errors for all three devices.

PREFACE

This final technical report was prepared by John R. Welch of Threshold Technology Inc. It was prepared under Contract Number F30602-76-C-0178 in accordance with Exhibit Line Item A003 for the Air Force Systems Command, Rome Air Development Center, Griffiss Air Force Base, New York. Capt. Robert A. Curtis (IRAD) was the Project Engineer.

The author wishes to acknowledge his gratitude to Dr. Bruno Beek and to Captain Curtis for the original conception of this study, to Dr. Thomas B. Martin and to Marvin B. Herscher for their guidance and support, to Phillips B. Scott for his leadership and engineering assistance and to Bruce Ball for his consultation and programming contributions.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

LIST OF ILLUSTRATIONS

LIST OF TABLES

EVALUATION

This study was done to evaluate various data entry processes. Experiments were run to evaluate voice data entry, keyboard entry, and Graf Pen entry for entry time and the error rates involved. By providing a statistical basis of actual human factor experiments, the future design of data entry systems can incorporate the results of this report to improve the overall data entry process.

Robert A. Curtis

ROBERT A. CURTIS, Captain, USAF
Project Engineer

SECTION I

INTRODUCTION

A. Objectives

The objective of this program has been to perform an analysis and an experimental evaluation of human factors and other problems associated with several methods of inputting data into an information data handling system. The input modes to be studied were to include voice and several other alternatives. Measurements were to be made of efficiency and accuracy, and an assessment was to be made of the devices' applicabilities to future man-machine interfaces.

B. Background

The electronic data processing (EDP) community has, in recent years, greatly expanded its overall capabilities. Today's EDP machines are faster, more reliable, cheaper and easier to maintain. Intelligence Data Handling Systems (IDHS) have also been significantly improved, primarily with respect to the handling of data once it is in the computer. For example, IDHS can correct spelling errors, add and subtract words, change paragraph positions, etc. However, little if any attention has been paid to the problem of capturing the data at its source. In fact, one of the major limitations of IDHS systems is in getting information into the computer. Recently, automatic speech recognition (ASR) systems have shown the potential of becoming useful means of data entry and control. In fact, several limited vocabulary, discrete word recognition systems are already being marketed. Other recognition devices with different capabilities are now being developed and should also be available in the near future. These devices, although not natural language systems, may provide necessary data entry solutions for a large set of problems.

In order to apply voice or any other data entry device effectively it is necessary to obtain reliable experimental measures of the advantages, limitations, and the basic operating characteristics of the device. Voice, as an input mode, is so new that there have been few, if any, carefully controlled experiments to assess its capabilities.

Current successful voice data entry systems provide limited vocabulary, speaker dependent, isolated word recognition. With highly experienced operators, these systems are capable of achieving error rates of 1% or less for relatively large vocabularies. With inexperienced speakers the error rates may initially be as high as 3 or 4%, but this performance level is still useful and impressive. These voice input systems are currently operational in a large number of commercial and Government applications. Many of these applications involve data entry by personnel whose hands are occupied by other tasks. Other units have been sold to provide environmental control and mobility for disabled individuals who have no use of their hands or legs. These applications for voice input are very effective since voice is really the only data entry alternative that can match the efficiency of manual data input without requiring use of the hands.

The question is, however, can a speech recognition system, which recognizes a limited vocabulary of isolated words and which makes recognition errors, compete with more conventional data entry systems in realistic problem settings in which hand occupation is not so complete as to eliminate manual data entry from consideration.

If hand occupation is not a dominant factor then speech input still has several advantages. Ochsman and Chapanis[1] have demonstrated that natural language voice communications are superior to non-voice communication modes for cooperative problem solving by humans. It is also clear that most humans are capable of speaking at data rates that exceed the transcription speeds of all but the most highly skilled stenographers. Voice, furthermore, provides eyes-free data entry without requiring extensive training as in the case of, for example, touch typing.

One good example of a successful commercial application for voice input which neither relies on hand occupation, nor is a high volume data input situation, is in voice programming of numerically controlled machine tools (VNC). In this application the operator must use his eyes and his mind extensively while entering data. The success of the system depends primarily on the computer program which converts the voice-entered dimensional data into a program for controlling the machine tool. Voice input augments this success, however, by providing freedom of the eyes and mind and a degree of naturalness that greatly enhances the effectiveness of the man-machine interface.

C. Summary of Work Accomplished

The experiments described in this report have been designed to provide more information about the inner workings of data entry systems employing voice, keyboard, and a graphical menu entry device. There are a number of factors which affect performance in such a system, in addition to the entry device itself. Some of the principal factors are:

1. Problem Setting
2. Data type
3. Prompting structure
4. Feedback mode
5. Degree of hand occupation
6. Operator experience

[1] Ochsman, R.B. and Chapanis, A., "The Effects of 10 Communication Modes on the Behavior of Teams During Cooperative Problem-Solving," Int. J. Man-Machine Studies, Vol. 6, 1974, pp. 579-619.

Because of time limitations and because of the preliminary nature of these tests we have not included extensive training or fatigue as factors in the experiments. As a result, the experiments reflect performance levels and problems that would be encountered by relatively infrequent users of data entry systems. The resultant entry speeds and error rates do not reflect levels attainable by highly skilled operators. This may seem like an unfortunate choice from the point of view of high volume applications but in practice there are probably more potential data entry applications of this variety than of the high volume variety. In addition, the ultimate success of voice input will probably be greatest in such applications simply because it is potentially the most natural way for inexperienced users to communicate efficiently with computers.

Two data entry experiments were performed. In both experiments, comparisons were made of speed and accuracy for voice, keyboard and Graf Pen entry devices. Both experiments included tests with and without voice response feedback, and with and without hand occupation. The first experiment, which we call the High Speed Data Entry (HSDE) test, is a measure of entry performance in copying single strings of numeric and alphanumeric data. The second experiment, which we call the High Complexity Data Entry (HCDE) test, is a measure of performance in entering simulated flight data control messages. In this experiment, the subject's ability to interpret an English language statement and convert it to a series of data entry fields had as much effect upon performance as did the raw speed of the data entry system.

Section II of this report describes both experiments in full. Section III presents the analysis of the experimental data. Section IV provides a summary and discussion of the experimental results.

Hopefully the data which has been obtained by running these experiments will be useful in guiding the design of a wide range of data entry systems. It must be kept in mind, however, that the results reported here are relative to very specific equipment configurations. In some cases, the experimental setups do not reflect optimum usage of the entry devices. Voice input, for example, was used somewhat suboptimally because it was not trained with as many repetitions as would be used with professional users, nor was the operator training itself anywhere near as extensive, because of time limitations when running the tests. In addition, after running the tests, it was discovered that an interrupt priority error had been adding unnecessary variable delays of up to 100 milliseconds to the voice input response time. These delays directly reduced voice entry time and indirectly affected voice entry accuracy and time by making it difficult for the subjects to establish a consistent entry rhythm. The graphical input device, as another example, could have provided a higher level of performance than it did if it had been configured as a light pen. Such configuration was not feasible, however, within the limitations of the testing budget. Finally, the voice response unit used for feedback would have performed more favorably if it had had a faster speaking rate and a larger vocabulary.

In spite of these problems a great deal of, sometimes surprising, information has been obtained which should generalize to other situations and which should help to guide future research.

Section II

DESCRIPTION OF TESTS

A. Selection of Data Entry Variables to be Tested

The data entry tests which have been run were selected from a very large set of possibilities. In this section, we will discuss some of the dimensions of the data entry problem and will indicate why the particular test configurations were chosen.

1. Data Entry Scenarios

Two separate data entry tests were performed. The first was a test of a relatively simple data copying task such as sorting, program keying, or general bulk data entry, for which the problem is to find the most efficient way to enter large quantities of data. The second test was a simulation of a complex, highly structured task such as flight traffic data entry or programming of numerically controlled machine tools. In this kind of system, the problem is to maximize the convenience and the comprehensibility of the system so as to ensure accuracy and to support the user's thought processes.

A data copying system is usually characterized by a relatively simple control program with a fixed data entry vocabulary. Such systems require little or no prompting but must have very rapid feedback and an efficient error correction mechanism. A complex data entry system, on the other hand, usually has a sequential hierarchical structure with different data entry vocabularies at each stage of the hierarchy. Prompting is critical in such systems, particularly for inexperienced users. There is a tradeoff, however, between prompting verbosity and entry speed. Inexperienced users require more detailed prompts. Gains and Facey[1] recommend that the user be given a simple method for selecting the degree of verbosity of the prompts. Prompting in this kind of system also provides a feedback function for entries which control branching. If the prompt is received for the branch which he selected, the user knows that the system properly interpreted his request.

In the simple data copying experiment, we have simulated a system which is typical of the state-of-the-art for such data entry configurations. In the complex data entry experiment, the system which was simulated was not optimized for ease of use and fell far short of the human factors standards recommended for such systems by Gaines and Facey[1] and Kennedy[2]. The system

---

[1] Gaines, B.R. and Facey, P.V., "Some Experience in Interactive Systems Development and Application", Proc. IEEE. Vol. 63, No. 6, pp 894-911, June 1975.

[2] Kennedy, T.C.S., "The Design of Interactive Procedures for Man-Machine Communication", Int. J. Man-Machine Studies, Vol. 6, pp 309-334, 1974.

was, in fact, set up as if it were to be used by relatively experienced subjects. The reason for this was primarily that limitations on the voice response unit vocabulary did not allow for a sophisticated prompting structure that was capable of automatically providing most of the users' training. Therefore, the tests were limited to comparing speed and accuracy of the entry devices and in comparing helpfulness of the prompting media after some degree of user training had been administered.

2. Size and Type of Vocabulary

Let us now consider how size and type of vocabulary can be expected to affect the data entry system performance for entry of individual fields. Entry of multiple fields does not change the effects of size and type of vocabulary except for menu oriented entry systems, which may become impractical because of the requirement for changing menus.

A large vocabulary (greater than 100 items) generally implies a vocabulary of words. Keyboard is a clear choice for such vocabularies because the words can be spelled either entirely or partially. Menu data entry is not advantageous in this case because of the excessive time required to scan such a large menu. Voice data entry would be ideal for large vocabularies of words if its speed and accuracy were suitably high, since an operator with no special skills could enter the equivalent of numerous keystrokes in a single utterance.

Medium size vocabularies (greater than 30 but fewer than 100 items) generally fall into two type categories, words or alphanumerics. Entry of a medium sized vocabulary of words can often be accomplished very effectively by voice or menu systems. The vocabulary size does not exceed the practical limitations of commercial voice data entry, nor does it result in a particularly impractical size of menu, except possibly near the upper limits of this vocabulary size range. Keyboard also can do a good job in this size range, but it has the disadvantage of requiring multiple keystrokes. The use of abbreviations can reduce the number of keystrokes to no greater than two per entry, but memorization of abbreviations increases the training requirements.

If the medium sized vocabulary is explicitly limited to alphanumerics either singly or as code strings, then keyboard no longer requires multiple keystrokes per entry. Menu data entry then has a disadvantage as compared to keyboard since, at best, it lets the operator proceed like a "one-fingered" typist. Voice data entry also has the disadvantage of requiring use of some form of phonetic alphabet. Some experience is required before an operator can memorize and master the use of a phonetic alphabet.

Small vocabularies (fewer than 30 words) can take the form of words, alphanumerics, or numerics. In this size range, vocabularies of words can be recognized very accurately by voice and the menu size is quite manageable for menu oriented systems. Keyboard still has the disadvantage of requiring multiple keystrokes or memorization of abbreviations.

Small vocabularies of strictly alphanumeric data once again tend to favor keyboard input. Menu input still has the disadvantage of being like one-fingered typing, but this is compensated to some extent by the fact that unlike a standard keyboard, the menu can be reduced in size and tailored exactly to match the vocabulary. Voice input again has the disadvantage of requiring use of a phonetic alphabet.

Numeric-only vocabularies can be processed very rapidly by special numeric keypads or by the numeric row of keys on a standard teletypewriter. It is possible to learn to touch-type such numeric keyboards with relatively little training. Furthermore, such keyboards can be used one-handed in applications requiring use of the other hand. Menu data entry by contrast cannot compare to keyboard because of its "one-fingered" nature. Voice entry, likewise, cannot compete with the speed of a numeric keyboard unless entire data strings can be entered by continuous speech. Continuous speech recognition systems are being developed, but they tend to be very costly and generally provide poor recognition accuracy compared to isolated word recognition systems.

From this set of possibilities for size and type of vocabulary, we have chosen to test small numeric and small alphanumeric vocabularies in the high speed data entry test, and small vocabularies of words, alphanumerics, and numerics sequentially selected from an overall medium sized vocabulary in the high complexity data entry test. As far as the subjects were concerned, the vocabulary in the high complexity test was a medium sized vocabulary of words and alphanumerics. The menu was medium sized and keyboard entry of words required two keystrokes.

3. Length of Data Fields

Length of data fields is a parameter which is applicable to alphanumerically coded fields or strictly numeric fields. If each field must be verified by a word such as "ENTER" or a carriage return, then short fields require more verification time per character than do long fields. On the other hand, time for entry of long fields is increased by the requirement that the operator mentally break the fields into smaller more easily memorized segments, which are then entered separately. By making field length a variable in the high speed data entry experiment, we have attempted to determine which of these two effects is dominant.

4. Hand and Eye Occupation

Occupation of the hands by some external task is a factor which strongly favors voice input. In the experimental designs, we have tried to provide a measure of the effect of hand occupation on data entry performance for some very specific hand occupation tasks. We have only considered situations in which both hands are occupied. For numeric-only data entry with keyboard, occupation of only one hand would be expected to have less effect than occupation of both hands. Likewise, occupation of one hand could have a reduced effect on entry by menu entry devices since these typically require use of only one hand.

Occupation of eyes favors voice or touch typing as input modes, but was omitted as a variable in the experiments because very little data entry is done completely from touch, sound, or memory without any visual input from notes, diagrams, lists or flowcharts. Consequently, all tests were performed with a requirement to use the eyes to read the data to be entered.

5. Feedback and Prompting

In the high speed data entry scenario, prompting is not a requirement, but feedback is important. Therefore, we provided a comparison of visual and voice response feedback in the HSDE tests. Visual feedback was presented on a CRT since that is the way it normally would be presented for keyboard entry and since that enabled problem presentation (from a Burroughs Self-Scan Display) and feedback to be widely separated from each other in physical location.

Voice response feedback was never provided in lieu of visual feedback, but was only used to augment visual feedback. There are two reasons for this. One was that visual feedback would almost always be present with keyboard input anyway, and the second was that voice response feedback does not have the retention qualities required for correction of long character strings or recovery from lapses in the user's attention.

Voice response feedback could have been provided either after each character or at the end of complete data strings. We chose the former because it seemed preferable for correction of character errors as they were made and because we believed it would be faster. In retrospect, we believe that there would be value in performing further tests involving rapid voice response feedback at the end of each string.

In the high complexity data entry tests, feedback per se was always provided via the CRT. Voice response was used only to augment the identical prompting messages that were simultaneously provided on the CRT. Voice response was not used by itself because it had no retention capability. Once issued, prompts could not be reissued (except perhaps by provision of a special repeat-prompt command).

B. The High Speed Data Entry Tests

The High Speed Data Entry (HSDE) tests were comparisons of speed and accuracy for entering strings of randomly selected numbers and letters which were presented on a 16 character wide, one centimeter high Burroughs Self-Scan Display. The variables of the experiment were the entry mode (device), the type of data characters (numeric and/or alphabetic), the length of the data strings, the presence or absence of an external hand occupation task, the type of feedback, and the test repetition number (trial). The subjects were randomly selected from a pool of subjects with a wide range of experiience levels with the devices tested. The amount of pre-test training was limited, so that for these particular tests all subjects could be considered novices.

1. Experimental Design

The high speed data entry tests employed a factorial design to investigate all combinations of the basic factors (variables).

A factorial design is argued by Fisher[1] to have the following advantages over testing each factor individually:

a. Greater efficiency - each individual factor is evaluated with as much precision as if the entire experiment were devoted to that factor alone;

b. Greater Comprehensiveness - in addition to the effects of single factors, all their possible combinations are evaluated;

c. Wider inductive basis for drawing conclusions, since variables are not treated in isolation.

The factors investigated, the number of levels per factor and the descriptions of the levels are given in Table 1.

TABLE 1

HSDE FACTORS

	FACTOR NAME	NUMBER OF LEVELS	DESCRIPTION OF LEVELS
(a)	Entry Mode	3	Voice, Keyboard, Graf Pen
(b)	Data Alphabet	2	Numeric Only, Alphanumeric
(c)	Data Length	2	3 Characters, 10 Characters
(d)	Hand Occupation	2	Pushbutton Required, Not Required
(e)	Feedback	2	Visual and Voice Response, Visual Only
(f)	Trial	3	Trial 1, Trial 2, Trial 3

This factorial design involved 48 subjects and 144 tests, since the experiment was run without replication.

In this set of experiments, systematic bias due to individual subject variations and order of presentation of tests was eliminated by randomizing both factors. It would have been preferable to group subjects according to ability levels and to introduce that classification as an additional factor in the experiments, but there were not enough subjects available with known experience levels to make that feasible in this test.

---

1 R.A. Fisher, "The Design of Experiments", pp 93-108, Hafner, New York, 1971.

The statistical analysis of these tests involved comparisons of all the main factors and of interactions between factors for the following measurements:

a. Average Time Per Correct Character

b. Percent Wrong Characters Before Correction

c. Percent Wrong Characters After Correction

d. Percent Wrong Character Strings After Correction

2. Hardware Configuration

The test set-up is diagrammed in Figure 1. The test system was controlled by a Data General Nova 800 computer which was part of the VIP-100 voice recognition system. Test data was displayed on a Burroughs 16 Character Self Scan. Data entry was performed via the VIP-100 voice input system, the Science Accessories Corporation Graf Pen, or a Lear Siegler ADM-3A CRT keyboard. Immediate feedback of the entered data was displayed on the CRT for all three data entry modes. Feedback messages were also provided in parallel by a Speech Technology Corporation voice response unit for those tests requiring voice feedback. A Teletype was used for controlling the experiments, producing a hard copy of the test data and the subjects' responses, and printing time and error rate statistics at the end of each test. A set of pushbuttons separated by 14 inches was provided to force the subject to use both hands simultaneously to obtain a new string of characters on the prompting display.

a. The VIP-100 Voice Recognition System

The VIP-100 voice recognition system is an isolated word recognition system which is normally trained to recognize a specific vocabulary spoken by a particular person. In the HSDE tests, the vocabulary, which is listed in Table 2, consisted of numbers, code words representing letters and several control words. In these tests, five repetitions were used to train each word, and if recurring recognition errors were encountered during the tests, particular words were retrained with five new repetitions. Somewhat better recognition results would have been obtained by using 10 training repetitions per word as is normally done with the VIP-100 system. Five repetitions were used, however, to keep the overall subject preparation effort for voice input at a level commensurate with the other input system.

Once trained, the VIP-100 system responds to the spoken voice much as a keyboard responds to depressing of keys. The primary difference is that the VIP-100 system does sometimes make errors in recognition. It also sometimes cannot classify a sound as belonging to the expected set of words. In that case, it provides a reject indication by flashing a red light and by providing any other indication desired under software control. Finally, it has inherent entry speed limitations due to a requirement to leave at least 100 milliseconds of silence between successive words being entered, and in this test configuration, inadvertently had unnecessary additional delays of

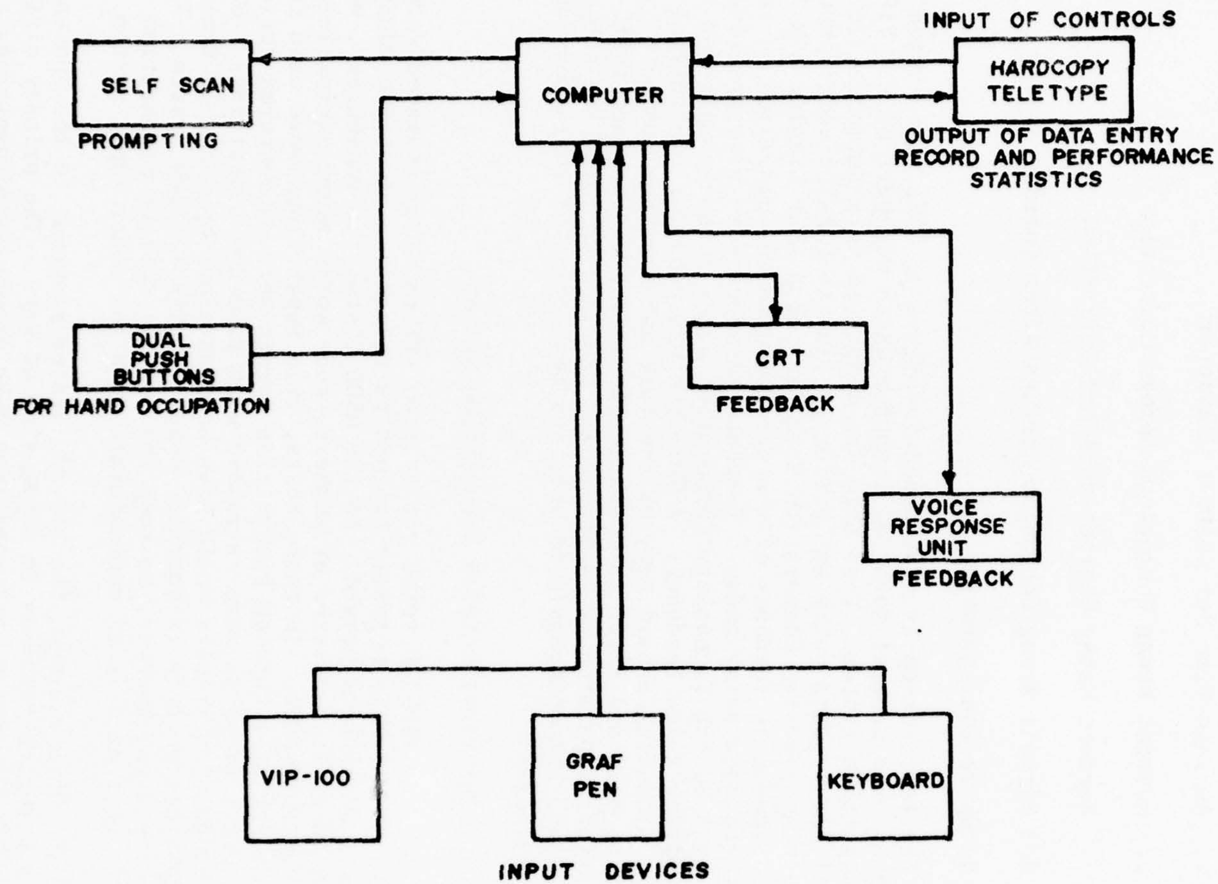


Fig. 1 Hardware Configuration for Data Entry Experiments

up to 100 milliseconds due to errors in the priority interrupt structure.

TABLE 2

HSDE RECOGNITION VOCABULARY

NUMBER	TTY CHARACTER	VOICE ENTRY AND FEEDBACK WORD
0	0	ZERO
1	1	ONE
2	2	TWO
3	3	THREE
4	4	FOUR
5	5	FIVE
6	6	SIX
7	7	SEVEN
8	8	EIGHT
9	9	NINE
10	RUB	BACKSPACE
11	SHIFT-RUB	DELETE
12	CR	ENTER
13	A	ADDRESS
14	B	BLOCK
15	C	CODE
16	D	DATA
17	E	END
18	I	IDENTIFICATION
19	L	LOCATION
20	M	MESSAGE
21	N	NAME
22	R	READY
23	S	SYMBOL
24	T	TIME
25	U	UNIT

For more technical details on the VIP-100 system, see Appendix B.

b. The Graf Pen

The Science Accessories Corporation Graf Pen is a graphical input device which converts arrival times of sonic pulses into distance measurements. The pulses are generated by a hand held spark generator (stylus) and detected by a pair of microphones mounted at the edges of the working digitization surface. In these tests, the Graf Pen was interfaced with the Data General computer and a program was written to provide data entry from a fixed data "menu". Grid coordinates were converted to menu selections by means of a simple table-look-up. Point sensor microphones were used in the tests because of mechanical problems with the available line sensor microphones, and this required the use of large radius circular boundaries for the menu grid coordinates instead of rectangular coordinates. Initial alignment of the menu grid pattern and the Graf Pen microphones was accomplished by means of an interactive alignment and calibration program.

Data entry via the Graf Pen required that the subject locate the desired block on the menu grid pattern and then press the stylus tip down somewhere within the block. There is one anomaly with the Graf Pen. The stylus tip is offset by about 3/32 of an inch with respect to the spark gap, so that it was possible to press the stylus down on the correct block, but obtain the coordinates for a neighboring block. The subjects were told of this problem and were requested to keep the stylus center and hence, the spark gap within the desired block. Nevertheless, there were occasional errors and syntactical rejects which resulted from missing the correct block because of the offset.

For further technical details on the Graf Pen, see Appendix C.

c. The LSI-ADM-3A CRT Terminal

The Lear Siegler ADM-3A is a single unit keyboard and CRT terminal. The keyboard is designed to teletypewriter layout. The display was set up for twenty-four 80-character lines on a 12 inch screen. Data entry was from the bottom of the screen with upward page scroll.

d. The Speech Technology Corporation M-200 Voice Response Unit

The STC M-200 is a fixed vocabulary voice response unit which sounds fairly natural because it provides (highly compressed) reproduction of actual human speech. The VRU is essentially a synthesizer for a formant tracking vocoder. The vocoder analysis is performed off line (at STC) to provide digitization and compression to about 600 bits per second. The digital words are stored in ROM in the M-200 unit. Vocabulary for the M-200 is custom ordered from STC who burns it into ROM. Words can be selected from a reasonably large and growing vocabulary list for virtually no charge or they can be recorded and digitized to order for $150 per second of speech.

The VRU was interfaced to the Nova-800 computer and programs were written to select particular words or sequences of words comprising messages.

The VRU word list is given in Table 3. The VRU performed flawlessly, but as will be discussed later, both its limited vocabulary and its relatively leisurely speaking rate adversely affected its performance in the data entry tests. For more technical details on the STC voice response unit, see Appendix D.

TABLE 3

VOICE RESPONSE UNIT VOCABULARY

1.	NAME	21.	FIVE
2.	IDENTIFICATION	22.	SIX
3.	ADDRESS	23.	SEVEN
4.	LOCATION	24.	EIGHT
5.	TIME	25.	NINE
6.	ENTER	26.	ERROR
7.	TYPE	27.	DELETE
8.	OF	28.	BACKSPACE
9.	MESSAGE	29.	READY
10.	UNIT	30.	RELEASE
11.	NUMBER	31.	NOT
12.	DATA	32.	CHANNEL
13.	CODE	33.	IN
14.	DEVICE	34.	OUT
15.	BLOCK	35.	SYMBOL
16.	ZERO	36.	START
17.	ONE	37.	STOP
18.	TWO	38.	LINE
19.	THREE	29.	END
20.	FOUR		

3. The Test Program

The high speed data entry test program generated random strings of digits or mixtures of digits and alphabetic characters either in response to the hand occupation pushbuttons or automatically upon completion of each data entry trial. Elapsed time was measured by the program using the computer's real-time clock, and, at the end of a sequence of tests, average entry rates and error rates were computed and displayed. The program was set up so that data entry input could be selected to be from voice, keyboard or Graf Pen.

Test data displayed to the subject was always via Burroughs Self-Scan, but recognition feedback was presented either by CRT, or by CRT and Voice Response Unit simultaneously.

a. Recognition Vocabulary and Graf Pen Menu Layout

In the voice mode, there were two different recognition vocabularies:

1) The digits 0-9
2) The digits 0-9 and the set of 13 words listed in Table 2 as entries 13 through 25.

In addition, the words "BACKSPACE", "DELETE", and "ENTER" were recognized at the appropriate time to produce erasure of the last word entered, deletion of the entire entry, and entry of the complete string of words, respectively.

The choice of alphabetic characters and corresponding voice entry and feedback words was dictated by the vocabulary available in the voice response unit. That vocabulary was limited and was chosen primarily for the high complexity data entry tests. One disadvantage of voice entry of alphabetic characters is that the person doing the entering must memorize a phonetic alphabet; direct entry of the letters does not work as well. In these tests, the particular phonetic alphabet was dictated by the voice response unit. In the keyboard and Graf Pen modes, the vocabularies were exactly the same as for voice except that only the first letters of the alphabet words were entered. Graf Pen entries were selected from a menu like that shown in Figure 2.

b. Length of Data Strings

The number of characters in a string was a variable that could be set at the beginning of each test. The numbers 3 and 10 were used in the tests.

c. Hand Occupation

Hand occupation was selected as an option at the beginning of each test. Hand occupation was implemented by requiring that the subject simultaneously push two buttons separated by 14 inches in order to generate each new test data string. In the high speed data entry tests, only an instantaneous push was required, so that the hands were usually occupied for less than a second per data string.

d. Timing and Number of Samples Selection

A counter was provided to control the number of samples to be presented in each test. The number to be presented was selected at the beginning of each test. Ten samples were used in the 10 character tests, and twenty-five samples were used in the 3 character tests. The data generation and timing programs were started in response to the control prompt "HIT CR TO START".

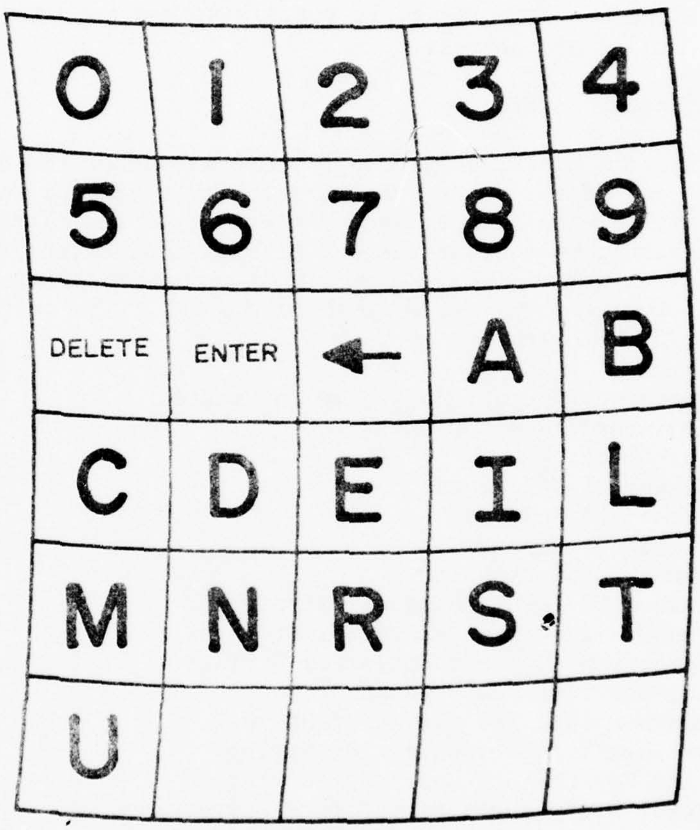


Fig. 2 Graf Pen Menu for HSDE Tests

The test data generation and timing stopped automatically when the specified number of data strings was presented and entered.

Three timing components were measured:

1) Actual data entry time, i.e., accumulation of all times between completions of prompts and completion of data entry.

2) Lost time, i.e., time required for generation of test data strings.

3) Time for retraining, if retraining was required in the middle of the test.

e. Measurement System

During the tests, a hardcopy record was generated of the test data (the prompt) and of the subject's character-by-character entries, including backspaces, deletions, and system rejects. In addition, a battery of counters were operating so that at the end of each test extensive performance statistics could be computed. These statistics were then printed on the off-line Teletype and displayed to the subject on the CRT. The statistics that were measured are the following:

. Encoding (Operational) Time In Minutes
. Training Time In Minutes
. Lost Time
. Number Of Utterances
. Number Of Rejects
. Number Of Erasures
. Number Of Backspaces
. Number Of Total Character Strings
. Number of Correct Character Strings
. Percent Of Correct Character Strings
. Average Time Per Utterance
. Minimum Time Per Character String
. Average Time Per Character String
. Time Per Correct Character String
. Total Wrong Characters Before Corrections
. Percent Wrong Characters Before Corrections
. Total Correct Characters
. Percent Of Correct Characters
. Percent Correct Characters (Per Utterance)
. Average Time Per Correct Character
. Variance of Encode Time For Character Strings
. Std. Deviation Of Encode Time For Character Strings

4. Subject Selection

All of the subjects were employees or family members of employees of Threshold Technology Inc. (TTI), except for one subject, who was a customer, another who was a supplier and a third who was a consultant for a customer working at TTI. The order of performing the tests was selected at random and the subjects were selected alphabetically with numerous deviations from alphabetical order dictated by availability. Generally, however, the assignment of subjects to test conditions was randomized.

This random assignment resulted in a distribution of subjects among the three entry modes as shown in Table 4.

TABLE 4

EXPERIENCE DISTRIBUTION OF SUBJECTS IN HSDE TESTS

MODE	NO - EXPERIENCE	LITTLE EXPERIENCE	HIGHLY EXPERIENCED	EXPERT
VOICE	7	7	2	0
KEYBOARD	1	1	12	2
GRAF PEN	12	4	0	0

In this table, expert means a person who has been professionally employed to enter data by that particular device. The two expert typists were members of the secretarial staff of TTI. Highly experienced means a person who has spent many hours entering data by the device. The two highly experienced voice operators are engineers at TTI who have worked with voice input many hours. The twelve highly experienced keyboard operators are either computer programmers or fair non-touch typists. Little experience means a person who has used the device for a total of no more than one or two hours. No experience means a person who has not used the device at all prior to these tests.

From this breakdown it is clear that keyboard was given a distinct advantage in these tests by virtue of prior experience of the keyboard subjects. Voice entry had many fewer experienced operators and all of the Graf Pen tests were performed by operators with little or no experience. While this subject distribution is biased toward keyboard in terms of experience breakdowns, it is probably an accurate reflection of the experience levels of skilled white collar workers, except for the fact that it has a higher incidence of voice data entry and Graf Pen experience than would be expected in a more typical subject cross section.

Most subjects were used only with one test condition. Four subjects were used twice with different input modes. One subject was used with three different input modes. There did not seem to be a great deal of generalized learning in this test so that training was not expected to carry over from one input mode to another. Nonetheless, multiple use of subject was avoided except near the end of the tests when all available subjects had already been tested.

5. Instructions to Subjects

Each subject was first told that the purpose of the experiment was to compare the data entry speed and accuracy of three input devices; voice, Graf Pen and keyboard. He was then told which device he would be working with and was given a description of how the test would be run. He was told what kind of characters would appear on the Self-Scan, how many characters per string, how many strings per test, and how many tests there would be and whether the feedback would be by CRT or by CRT and voice response.

He was then given a description of how to operate the entry device. In the case of keyboard, all that was necessary was to indicate which keys were used for correction of a single character, for deletion of the entire entry and for final verification of the entered string. It was also necessary to explain how to respond to rejects.

For Graf Pen operation, the description included an explanation of how the Graf Pen worked, i.e., by sonic pulses, etc. This led to the precaution not to place anything between the stylus and the microphones and to the precaution to keep the stylus spark gap within the entry grid block. The Graf Pen description also included an explanation of the backspace, delete, enter and reject control functions.

For voice input, the orientation procedure was much more complicated. It was necessary to explain how to wear the head-mounted microphone, to set the volume control to match the subject's speaking level, to explain how the voice input system would be trained, to explain the use of the belt box microphone switch and to give instructions for how to speak to the system. The instructions included recommendations for pausing between each word, speaking in a relatively short, clipped manner, and never stretching out a word that was misrecognized to allow the VIP-100 to "hear" better.

Training required five repetitions of each word of the 26 word vocabulary and usually did not take more than about four minutes. In most cases, at least one word was immediately retrained, however, either because the subject had spoken an erroneous word during training or because he had lost track of the training repetition count.

Use of the hand occupation pushbuttons was generally one of the last things explained.

Finally, the subjects were instructed to strive for maximum possible speed consistent with reasonable input accuracy.

Many further instructions were usually required during the first test repetition. The subject was told during the first test not to worry about time, since it was generally to be used as a training run. As a result, the timing data from the first trial was highly erratic. The kinds of problems which were usually encountered during the first trial were confusion about how to handle rejects, incorrect use of the backspace and delete command, and, in the case of voice, recognition problems both with the data and with the correction commands.

C. High Complexity Data Entry Tests

The high complexity data entry tests were comparisons of speed and accuracy for entering simulated flight data control messages. The messages were typed on a problem sheet in the form of English sentences describing the data fields to be entered. All fields were underlined for clarity, and entry of the data fields was prompted by the data entry system. In order to test the efficiency of prompting, some confusion was left in the problem statements by making the order of presentation of some of the data fields different from the order in which they were prompted.

1. The Experimental Design

The experiment had a factorial design. The factors, the number of levels per factor and the descriptions of the levels are given in Table 5. This factorial design involved twenty-four subjects and 72 tests. The tests were run without replication other than the three trials for a given subject.

TABLE 5

HCDE FACTORS

FACTOR NAME	NUMBER OF LEVELS	DESCRIPTION OF LEVELS
(a) Entry Mode	3	Voice, Keyboard, Graf Pen
(b) Prompting Mode	2	Visual and Voice Response, Visual Only
(c) Hand Occupation	2	Pushbutton Required, Not Required
(d) Subject Experience	2	No Experience, Some Experience
(e) Test Repetition	3	Trial 1, Trial 2, Trial 3

The entry modes were a VIP-100 with 16 channel preprocessor, a Lear Siegler CRT terminal, and a sonic Graf Pen used in the flat tablet menu mode.

Prompting was provided on the Lear Siegler CRT in all tests. For voice and visual prompting, messages from the Speech Technology Corporation voice response unit were added. The prompting messages were the same for both devices.

Hand occupation was simulated by a requirement for the subject to hold down two pushbuttons simultaneously for a total of 3.5 seconds per input message. The pushbutton requirement could be satisfied concurrently with the data entry process, so that for voice input, little or no additional time was required for hand occupation.

Several experience factors were considered in designing this experiment. First was the subjects' experience with the entry device. We have observed marked speed differences between subjects who know where the characters are on a keyboard, and subjects who must perform a visual scan to find every character. Likewise, with voice entry, there has been a consistent (though sometimes small) difference in performance between those who have had hours of experience talking to speech recognition equipment and those who have never entered data by voice. Therefore, for these tests, subjects were randomly selected from one of two categories; those with zero experience using the entry device, and those with slight to moderate experience. By making experience a specific factor in the experimental design, its effect can be measured and balanced out of the estimate of experimental error. This not only provides a measure of the effect of experience, but also increases the precision of the experiments for evaluating the other factors.

A second experience factor was specific to the high complexity data entry test itself. This was such a complicated data entry scenario to learn that it was important not to use any subject for more than one test configuration. The use of subjects more than once would have required an experimental design which balanced the sequential (training) effects so that they would not be confounded with the other experimental factors. In future studies, such designs may be desirable since they make more efficient use of the available subject population. For this test, however, there were enough subjects to provide one subject per condition. This not only has simplified the experimental design and its associated statistical analysis, but also has provided a wider subject base for generalization of the results than would have been provided by fewer subjects running multiple tests.

Training within the tests was handled by giving the subjects a preliminary short test for training purposes before running the actual test. Then a measure of relatively short term training effects was obtained by having each subject run three separate tests with the same data entry configuration.

A final experience factor in the experimental design relates to the experience of the individual administering the tests. Experience in conducting the tests could conceivably have had an effect on the test results. An attempt was made to eliminate a systematic bias of this type by randomizing the order of performing the tests.

2. Hardware Configuration

The hardware configuration differed from that described in Section II-B-2 for the high speed data entry tests only in the following two ways.

a. The Burrough's Self-Scan was not used. Visual prompting and visual feedback were displayed on the Lear Siegler ADM-3A CRT Terminal only.

b. The set of two pushbuttons had to be held down for a total of 3.5 seconds per message in order to satisfy the hand occupation requirement.

3. The Test Program

The HCDE test probram was based on a program that was written to evaluate voice entry for Enroute Flight Data Control. In operation, the system issues prompting messages and then waits for rigidly formatted messages to be entered in response. The first prompt issued is always "TYPE OF MESSAGE?". The first input expected by the system is one of the words specifying the type of message such as AMEND or HANDOFF. Recognition of an acceptable input then produces a two character code on the display unit, such as AM or HA followed by a carriage return, line feed, and a new prompt. In most cases, the next prompt is "ENTER IDENTIFICATION NUMBER?", and the next entry expected is a three digit flight identification number to identify the flight data file to which the message applies. The three digits are entered and displayed one by one and are followed by CR, LF and a new prompt. The next entry expected by the system depends upon the kind of message selected by the first entry. AMEND, for example, expects the name of a flight plan data field such as AIRCRAFT TYPE or LOCATION, followed by an entry or entries appropriate to that data field. At the end of the message, the system almost always designates the end by issuing the prompt "END OF MESSAGE?". If the subject is satisfied that the data has been properly entered, he responds with the word END. This terminates the entry of that message and causes the prompt to be issued for the next message. At any time during message entry, the commands BACKSPACE and ERASE can be used respectively to delete either the last word entered or all entries for the entire message.

Data entry input could be selected to be from voice, keyboard, or Graf Pen. The prompting messages were presented both to the CRT and to the voice response unit. When no voice prompting was required, the audio output of the VRU was turned off. The system automatically generated message sequence numbers which corresponded to numbers on the problem lists, and a hardcopy record was made of all entries, backspaces, erasures, and system rejects. Elapsed time was measured by the computer's real-time clock.

4. Diagram of Prompts and Data Entry Syntax

Figure 3 is a diagram which illustrates the prompts, the general recognition vocabulary and the recognition syntax for the HCDE test program. In this diagram, the rectangular blocks represent prompting messages. Solid circles and solid diamonds represent points at which the system stops and waits for data input. The solid circles correspond to data entry from a subset of the total vocabulary and, in the case of numerical data, involve a count of the number of digits. The solid diamonds represent data entry which results in branching based on the contents of the data. There were only four branch points in the program. These are the points where type of message, data block name, aircraft type, and an I.D. number without a length constraint are entered.

At any point where the system expects data input, except for the first branch point, the recognition system will accept a backspace or erase command. The backspace command eliminates the last entry that was made and backs the data entry system up to the previous entry point. If the previous entry point would normally have been accompanied by a prompt, the system will

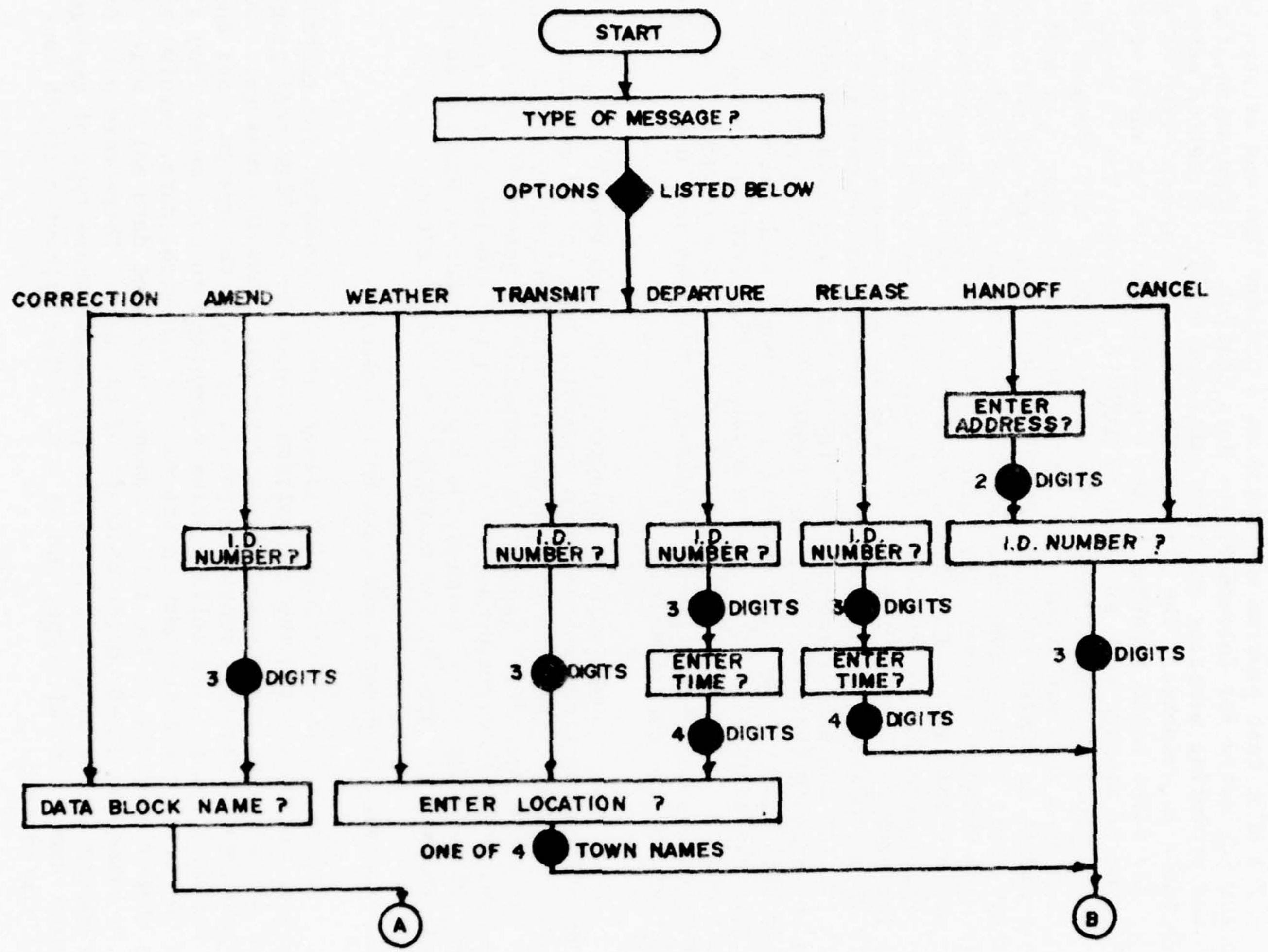


Fig. 3 Diagram of Prompting and Data Entry Structure (Sheet 1 of 2)

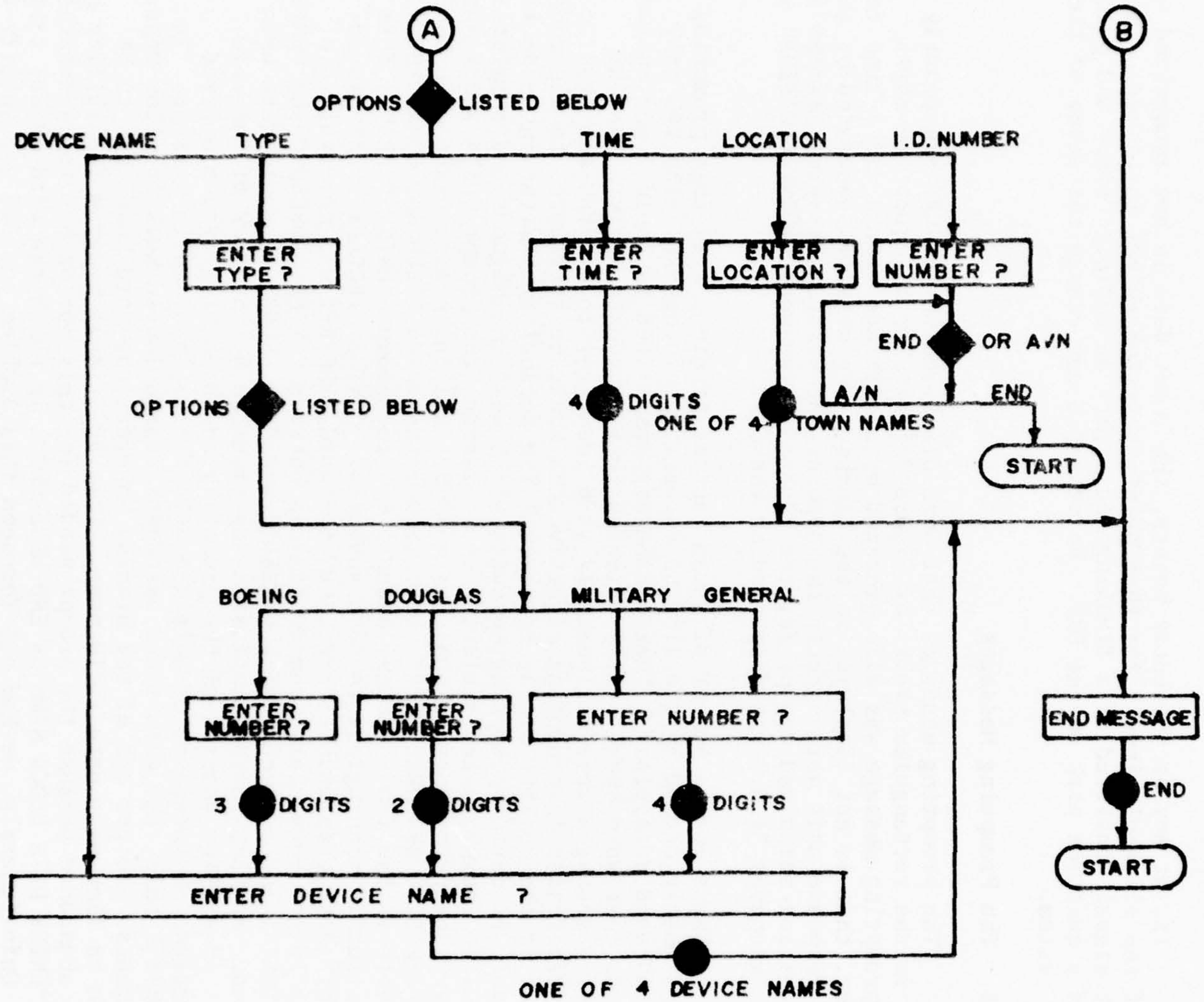


Fig. 3 Diagram of Prompting and Data Entry Structure (Sheet 2 of 2)

also reissue the prompt. The erase command backs the system up to the first branch point and issues the "TYPE OF MESSAGE?" prompt, but does not increment the message number counter.

If, at any of the entry points, the input data is not recognized as one of the allowable inputs for that point, a system reject is issued. The reject signal consisted of a flashing red light, an audible "beep" and print-out of a question mark on the CRT. Rejects did not change the state of the input system.

5. The Prompting Messages

The prompting messages that were displayed on the CRT are exactly as shown in the rectangular blocks of Figure 3 including the question marks. Each prompting message was also preceded by a carriage return and a line feed, however, that is not displayed in the blocks. The messages generated by the voice response unit were exactly the same as the written messages, except for the question marks and except for the fact that the word "IDENTIFICATION" was spoken whenever "I.D." was displayed on the CRT.

There were a number of human factors problems with this prompting system, mostly caused by the limited vocabulary of the VRU. The VRU only had a 30 word vocabulary. That vocabulary was listed in Table 3 and includes no words that were specifically oriented to flight data entry. The CRT prompting messages were deliberately made identical to the VRU messages so that differences in prompting effectiveness would be attributed to the media and not to the quality of the messages. The vocabulary limitations made it difficult to devise prompting messages which were unambiguous and completely helpful to inexperienced subjects. For example, the prompt "DATA BLOCK NAME?" was very confusing to most subjects. A more helpful prompt might have been "DATA FIELD TO BE CORRECTED?", but neither the words FIELD nor CORRECTED were available in the VRU. Another example was the prompt "ENTER TYPE?". This was virtually meaningless to most subjects. "ENTER AIRCRAFT TYPE?" would probably have been much clearer, but the word AIRCRAFT was not available. With a larger vocabulary, the confusions between "ENTER ADDRESS?" and "ENTER LOCATION?", and between "ENTER NUMBER?" and "I.D. NUMBER?" could have been reduced. Another serious confusion was produced by the lack of the word "OR". An earlier version of the prompting and branching system allowed corrections or amendment of multiple data fields in a single message. To accomplish this, the data field amendment branch looped back to allow repeated amendments. To get out of the branch, however, the word "END" had to be spoken in lieu of a data field name. This additional program flexibility had to be abandoned because the prompt needed for this option had to be something like "ENTER DATA BLOCK NAME OR END MESSAGE?", in which the word OR was critical. Unfortunately, we had not foreseen this problem and did not specify the word OR to be in the VRU vocabulary.

6. Recognition Vocabulary and Graf Pen Menu Layout.

The HCDE recognition vocabulary is listed in Table 6. There were 43 words in the vocabulary including the word STOP which is not used in the final system. The first column of Table 6 gives the vocabulary subsets for the words.

4

TABLE 6

HCDE TEST VOCABULARY

WORD NO.	VOCABULARY SUBSET	WORD	KEYBOARD CHARACTERS
0	NUMBERS	ZERO	0
1		ONE	1
2		TWO	2
3		THREE	3
4		FOUR	4
5		FIVE	5
6		SIX	6
7		SEVEN	7
8		EIGHT	8
9		NINE	9
10	CONTROL	BACKSPACE	RUB
11		END	RETURN
12		ERASE	SHIFT RUB
13		STOP	S
14	TYPE OF MESSAGE	AMEND	AM
15		CANCEL	CA
16		CORRECTION	CO
17		DEPARTURE	DE
18		HAND OFF	HA
19		RELEASE	RE
20		WEATHER	WE
21		TRANSMIT	TR
22	DATA BLOCK NAMES	TYPE	TY
23		DEVICE NAME	DE
24		LOCATION	LO
25		TIME	TI
26		I.D. NUMBER	ID
27	PHONETIC ALPAHBET	ALPHA	A
28		BRAVO	B
29		CHARLIE	C
30		DELTA	D
31	LOCATIONS	WILLIAMSPORT	WI
32		ALLENTOWN	AL
33		HAZELTON	HA
34		STILLWATER	ST
35	AIRCRAFT TYPE	BOEING	BO
36		DOUGLAS	DO
37		GENERAL	GE
38		MILITARY	MI
39	DEVICE NAMES	DISCRETE	DI
40		DME	DM
41		TRANSPONDER	TR
42		TACAN	TA

The second column lists the word that was displayed to prompt the subjects when training the voice recognition system. The third column lists the characters that were recognized by the keyboard entry version of the program. These same characters were provided as feedback in the voice and Graf Pen entry versions of the program, and except for the numbers, the letters, and the control characters, were always the first two letters of the entry. With keyboard data entry, the only feedback that was provided was an immediate echo of the entered characters. If the entered character, or pair of characters, was not acceptable to the keyboard entry system, a reject would be indicated by a question mark printed immediately following the characters. If the entry was acceptable to the keyboard system, that would be indicated by generation of the next prompting message.

Graf Pen entries were selected from a menu such as that shown in Figure 4, but with 5/8 inch spacing between grids. The characters used to label the menu locations were the same as used for keyboard entry. As can be seen in Figure 4, they are by no means self-explanatory, and greater clarity would have been achieved by spelling the words out in their entirety. The disadvantages with spelling them out would have been either that the lettering would have been very small or the menu would have to have been made larger. In retrospect, we believe that it would have been better to spell them out with small letters.

It would have been better still to program the Graf Pen to work in the light pen mode, for then the appropriate vocabulary subset could have been displayed for each stage of the entry process. This would not only reduce the search time for finding the proper menu locations, but would provide automatic prompting. In future studies, we recommend that a light pen implementation be tested and that a light pen or intelligent terminal type prompting structure be implemented for voice and keyboard input as well.

7. Entry Problems

Tables 7, 8, and 9 are the three sets of entry problems that were used in the tests. Each test consisted of 15 problems. Each problem was stated as a proper English sentence with a number of underlined data fields.

The three tests are not equal in the amount of data to be entered. In Table 10 we list the number of fields, words, and key strokes involved in the three tests. A data field is defined as an underlined entity in the problem statements of Tables 7, 8 and 9. Non-numeric fields always involve a single voice entry. Numeric fields require from two to five voice entries. A word is defined as a single voice entry, and the word count for each problem set includes the 15 words required to verify each entry. Keystrokes are counts of the number of individual keys including "RETURN" which have to be struck to enter the data by keyboard. This number is larger than the number of words because all non-numeric fields require two strokes for keyboard entry.

8. Subject Selection

All of the subjects were employees of Threshold Technology Inc.,

TYPE OF MESSAGE	AM	CA	CO	DE	HA	RE	WE	TR
DATA BLOCK NAME	TY	DE	LO	TI	ID			
AIRCRAFT TYPE	BO	DO	GE	MI				
DEVICE NAME	DI	DM	TR	TA				
LOCATION	WI	AL	HA	ST				
NUMBERS	0	1	2	3	4	5	6	7
A/N	8	9		A	B	C	D	S
CONTROL	←	ERASE	END					

Fig. 4 Graf Pen Menu for HCDE Tests

TABLE 7

HCDE TEST 1

1. CORRECT TYPE DATA BLOCK TO GENERAL 6969 WITH DEVICE NAME TACAN
2. TRANSMIT FLIGHT I.D. NO. 054 TO ALLENTOWN
3. TRANSMIT FLIGHT I.D. NO. 248 TO WILLIAMSPORT
4. AMEND TYPE DATA BLOCK OF FLIGHT I.D. NO. 219 TO BOEING 707 WITH DEVICE NAME DISCRETE
5. (REQUEST) WEATHER AT ALLENTOWN
6. CANCEL FLIGHT I.D. NO. 323
7. AMEND I.D. NUMBER DATA BLOCK OF FLIGHT I.D. NO. 092 TO C0655
8. AMEND LOCATION DATA BLOCK OF FLIGHT I.D. NO. 413 TO ALLENTOWN
9. CORRECT TIME DATA BLOCK TO 2719 HOURS
10. (REPORT) DEPARTURE OF FLIGHT I.D. NO. 482 FROM ALLENTOWN AT 0797 HOURS
11. CORRECT TYPE DATA BLOCK TO DOUGLAS 10 WITH DEVICE NAME D.M.E.
12. AMEND TYPE DATA BLOCK OF FLIGHT I.D. NO. 642 TO DOUGLAS 08 WITH DEVICE NAME TRANSPONDER
13. CORRECT I.D. NUMBER DATA BLOCK TO A972D
14. HANDOFF FLIGHT I.D. NO. 995 TO ADDRESS 07
15. HANDOFF FLIGHT I.D. NO. 225 TO ADDRESS 46

TABLE 8

HCDE TEST 2

1. CORRECT TYPE DATA BLOCK TO MILITARY 6873 WITH DEVICE NAME TRANSPONDER
2. (REQUEST) WEATHER AT STILLWATER
3. AMEND LOCATION DATA BLOCK OF FLIGHT I.D. NO. 857 TO WILLIAMSPORT
4. (REQUEST) WEATHER AT ALLENTOWN
5. HANDOFF FLIGHT I.D. NO. 330 TO ADDRESS 51
6. AMEND TYPE DATA BLOCK OF FLIGHT I.D. NO. 592 TO MILITARY 8745 WITH DEVICE NAME D.M.E.
7. (REPORT) DEPARTURE OF FLIGHT I.D. NO. 875 FROM HAZELTON AT 0342 HOURS
8. RELEASE FLIGHT I.D. NO. 635 AT 0248 HOURS
9. HANDOFF FLIGHT I.D. NO. 779 TO ADDRESS 10
10. TRANSMIT FLIGHT I.D. NO. 636 TO STILLWATER
11. AMEND DEVICE NAME DATA BLOCK OF FLIGHT I.D. NO. 060 TO TACAN
12. AMEND TYPE DATA BLOCK OF FLIGHT I.D. NO. 350 TO DOUGLAS 10 WITH DEVICE NAME D.M.E.
13. (REPORT) DEPARTURE OF FLIGHT I.D. NO. 526 FROM STILLWATER AT 1028 HOURS
14. (REPORT) DEPARTURE OF FLIGHT I.D. NO. 766 FROM WILLIAMSPORT AT 0883 HOURS
15. (REQUEST) WEATHER AT WILLIAMSPORT

TABLE 9

HCDE TEST 3

1. CANCEL FLIGHT I.D. NO. 241
2. RELEASE FLIGHT I.D. NO. 289 AT 1245 HOURS
3. (REPORT) DEPARTURE OF FLIGHT I.D. NO. 710 FROM WILLIAMSPORT AT 0817 HOURS
4. RELEASE FLIGHT I.D. NO. 854 AT 2138 HOURS
5. AMEND TYPE DATA BLOCK OF FLIGHT I.D. NO. 132 TO GENERAL 5774 WITH DEVICE NAME DISCRETE
6. CORRECT I.D. NUMBER DATA BLOCK TO C732B
7. TRANSMIT FLIGHT I.D. NO. 444 TO STILLWATER
8. (REQUEST) WEATHER AT HAZELTON
9. (REQUEST) WEATHER AT WILLIAMSPORT
10. TRANSMIT FLIGHT I.D. NO. 846 TO HAZELTON
11. HANDOFF FLIGHT I.D. NO. 255 TO ADDRESS 33
12. AMEND TYPE DATA BLOCK OF FLIGHT I.D. NO. 831 TO GENERAL 4739 WITH DEVICE NAME TACAN
13. RELEASE FLIGHT I.D. NO. 062 AT 2242 HOURS
14. CANCEL FLIGHT I.D. NO. 223
15. AMEND LOCATION DATA BLOCK OF FLIGHT I.D. NO. 072 TO HAZELTON

TABLE 10

DATA COUNTS FOR HCDE TESTS

TEST NO.	FIELDS	WORDS	KEYSTROKES
1	56	114	150
2	55	113	148
3	49	111	140

except for one who was a family member of an employee and another who was a technical visitor. Subjects were divided into two categories; those with some experience with the particular entry device, and those with no experience. The order of running the tests was randomized and the subjects were selected from a list of TTI employees in reverse alphabetical order. The first subject on the list was chosen who fit the experience category required by the particualr test that was to be run.

The experienced subjects for voice and keyboard entry were never experts. The voice enterers were engineers, salesmen, or programmers who have spent hours talking into a VIP-100 voice entry system but who have never been professional data enterers. The experienced keyboard subjects were all engineers or technicians who had spent many hours typing but who were not skilled touch typists. There was no shortage of subjects in these skill categories at TTI.

The inexperienced voice entry subjects were three relatively new employees of TTI who had never used a voice entry device before and one technical visitor. The inexperienced keyboard subjects were members of the production staff and one employee family member. Subjects in these two experience categories were not very numerous at TTI.

The experienced Graf Pen subjects were selected from the pool of subjects who had used the Graf Pen in the previous high speed data entry tests. The amount of experience gained from running the previous tests did not bring them up to the same relative experience levels as the experienced keyboard and voice entry subjects, but it is questionable whether any amount of experience with the Graf Pen and a different entry menu would have qualified them as experienced with the particular menu used in the HCDE tests.

9. Instructions to Subjects

Each subject was first told that the purpose of the experiment was to compare the data entry speed and accuracy of the three input devices. He was then told which device he would be working with and was given a description of how the test would be run.

It was explained how the system would prompt the subject and how at each stage of the data entry process the number of acceptable inputs would be a specific subset of the total vocabulary. A chart was provided on the wall in front of the subject which listed the set of entry responses appropriate to each prompting stage. It was pointed out that the only data to be entered were the underlined fields, but that the entry order was not necessarily the same as the order in which the fields were written into the problem statement. Particular emphasis was given to the AMEND and CORRECTION messages since both of those required that the name of the amended data block be specified in response to the DATA BLOCK NAME prompt. Almost nobody understood this until he had entered several messages of this type.

Each subject was then given a description of how to operate the entry device. In the case of keyboard, it was explained that all words were entered by typing their first two letters. It was also necessary to indicate which keys were used for correction of a single character, for deletion of the entire entry and for final verification of the entered string. He was also told how to respond to rejects.

In Graf Pen operation, the description included an explanation of how the Graf Pen worked, and precautions for its use. The Graf Pen description also included an explanation of the backspace, erase, enter and reject control functions.

For voice input, the orientation procedure was much more complicated. It was necessary to explain how to wear the head-mounted microphone, to set the volume control for proper speaking level, to explain how the voice input system would be trained, to explain the use of the belt-box microphone switch, and to give instructions on how to speak to the system. The instructions included the requirements for pausing between each word, speaking in a relatively short slipped manner, and never stretching out a word that was misrecognized to allow the VIP-100 to "hear" better.

Training required five repetitions of each word of the 43 word vocabulary and usually did not take more than about 10 minutes. In most cases, at least one word was immediately retrained, however, either because the subject had spoken an erroneous word during training or because he had lost track of the training repetition count. Somewhat better recognition results would have been obtained by using ten training repetitions per word as is normally done with the VIP-100 system. Five repetitions were used, however, to keep the overall subject preparation effort for voice input commensurate with the other input system.

Use of the hand occupation pushbuttons was generally one of the last things explained.

Finally, the subjects were instructed to strive for maximum possible accuracy consistent with reasonable input speed.

Many further instructions were usually required during the short training test. The primary problems encountered during the first test were generally related to the prompting and entry structures, and not to the use of the entry device. Most subjects were still having some difficulty interpreting the prompts well into the first of the three actual tests. There were, in addition, the usual confusions about how to handle rejects, and the backspace and erase commands, and, in the case of voice, there were recognition problems both with the data and with the correction commands. For experienced voice data entry subjects, the correction commands were a problem because they differed from the commands which they have used in other voice entry programs. These subjects were inclined to use their accustomed commands as a matter of reflex, and this resulted in numerous correction system errors.

Section III

RESULTS

A. Explanation of the Analysis of Variance Procedure Used in This Report

In Table 11 of Section III-B, for example, we summarize the analysis of variance for the Average Time Per Correct Character measurements. Column 1 of this table lists all of the individual factors, all interactions between two factors, and one significant interaction between three factors. The number of degrees of freedom (df) corresponding to the source of variation in column 1 is listed in column 2.

Column 3 displays the sum of squares for each factor or interaction.

The sum of squares totaled over all possible combinations of factors is given at the bottom of column 3.

In the two experiments discussed in this report there was no replication to provide an error estimate. In such a case we may derive an error estimate from the variance of the high order interaction terms. In the HSDE tests there were six factors. We expect that individual factors and interactions between pairs of factors may show statistical significance, but we have little reason to expect that 3, 4, 5 or 6 way interactions should generally be significant. Hence, the variance from these high order interactions can be used as a measure of uncontrolled variability in the test. To compute this measure we sum the "sum of squares" terms for all such interactions and then divide by the df's for these interactions. The result is the estimated mean square experimental error. In Table 11, this sum of squares and the mean square errors are tabulated in the next to the bottom row of columns 3 and 4.

If it happens that an occasional three or four way interaction is significant when compared to this mean square error, it means that by including that interaction in the measure we have overestimated the error. Usually the degree of overestimation is slight, and in any case it always leads to conservative estimates of statistical significance.

In the analysis of variance tables of this report, significance levels are tabulated in terms of probabilities that a deviation of that magnitude would not occur due to chance alone.

B. High Speed Data Entry Tests

The particular measurements analyzed are:

1. Average time per correct character
2. Percent wrong characters
3. Percent wrong character strings
4. Percent wrong characters before correction

The operational error rate or error rate after correction is analyzed with respect to both character errors and string errors. These two error measures are highly correlated. The primary difference between them is that the string error rate is usually higher than the corresponding character error rate and the length of the strings tends to be a significant experimental factor when evaluating string error rates. Error rate before correction is indicative of the basic error rate of the data entry device and the problem setting.

1. Entry Time Analysis

Table 11 summarizes the analysis of variance for the average time per correct character measurements. This analysis is performed for the data from the final two trials only, since using the first trial for training introduced very large time variations which are not necessarily related to actual speed of entry.

Alphabet, entry mode, and data length, in that order, can be seen to be the three most significant factors affecting entry time per correct character. The only other single factor which is significant is trial. There are three interactions between two factors which are significant at the .99 level or higher; length by mode, hand occupation by mode, and hand occupation by alphabet. Finally, there is one three-way interaction between hand occupation, length, and alphabet that is significant at the .99 level.

Figure 5 plots the entry speed as a function of the four individual factors which achieve statistical significance. It is particularly noteworthy that neither hand occupation nor feedback had a significant overall effect on entry speed.

The comparison of entry modes shows that keyboard was the fastest mode in these test, requiring an average of 29% less time per character than voice; and 22% less time than Graf Pen.

The alphabet comparison indicates that entry of numeric-only data requires about 25% less time than entry of mixtures of letters and numbers. This is not surprising since a smaller vocabulary reduces keyboard and Graf Pen scan time and reduces voice input error rates.

The data length comparisons show that 10-character strings required about 14% less time per character than 3 character strings. If the overhead required for verifying the old entry and requesting and reading a new entry were assumed to be equivalent to entering an additional character of data, the difference between 10 character strings and 3 character strings would be expected to be about 20%. The fact that the difference was less than 20% probably results from the requirement for rereading the 10 character strings several times in order to break them up into more easily memorized units.

The difference between trial two and trial three was only about 9%. This rather small increment indicates that the subjects had fairly well mastered the mechanics of the experiment by the beginning of trial two.

Figure 6 presents graphs of average time per correct character versus the interactions between entry mode and length and entry mode and hand occu-

TABLE 11

ANALYSIS OF VARIANCE OF
AVERAGE TIME PER CORRECT CHARACTER
TRIALS 2 AND 3 ONLY

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	FEEDBACK (F)	1	0.001	0.001	-	-
2.	HAND OCCUPATION (H)	1	0.013	0.013	-	-
3.	FxH	1	0.132	0.132	-	-
4.	LENGTH (L)	1	1.118	1.118	16.44	.999
5.	FxL	1	0.236	0.236	3.47	.90
6.	HxL	1	0.284	0.284	4.18	.95
7.	ALPAHBET (A)	1	4.208	4.208	61.88	.99999
8.	FxA	1	0.085	0.085	-	-
9.	HxA	1	0.493	0.493	7.25	.99
10.	LxA	1	0.010	0.010	-	-
11.	ENTRY MODE (M)	2	4.015	2.007	29.51	.9999
12.	FxM	2	0.172	0.086	-	-
13.	HxM	2	0.721	0.360	5.29	.99
14.	LxM	2	1.253	0.626	9.21	.999
15.	AxM	2	0.129	0.065	-	-
16.	TRIAL (T)	1	0.493	0.493	7.25	.99
17.	FxT	1	0.004	0.004	-	-
18.	HxT	1	0.005	0.005	-	-
19.	LxT	1	0.165	0.165	-	-
20.	AxT	1	0.023	0.023	-	-
21.	MxT	2	0.021	0.010		
22.	HxLxA	1	0.543	0.543	7.99	.99
	ALL INTERACTIONS BETWEEN 3, 4, 5 AND 6 FACTORS = ERROR	68	4.65	0.068		
	TOTAL	95	18.208			

GRAND MEAN = 1.489

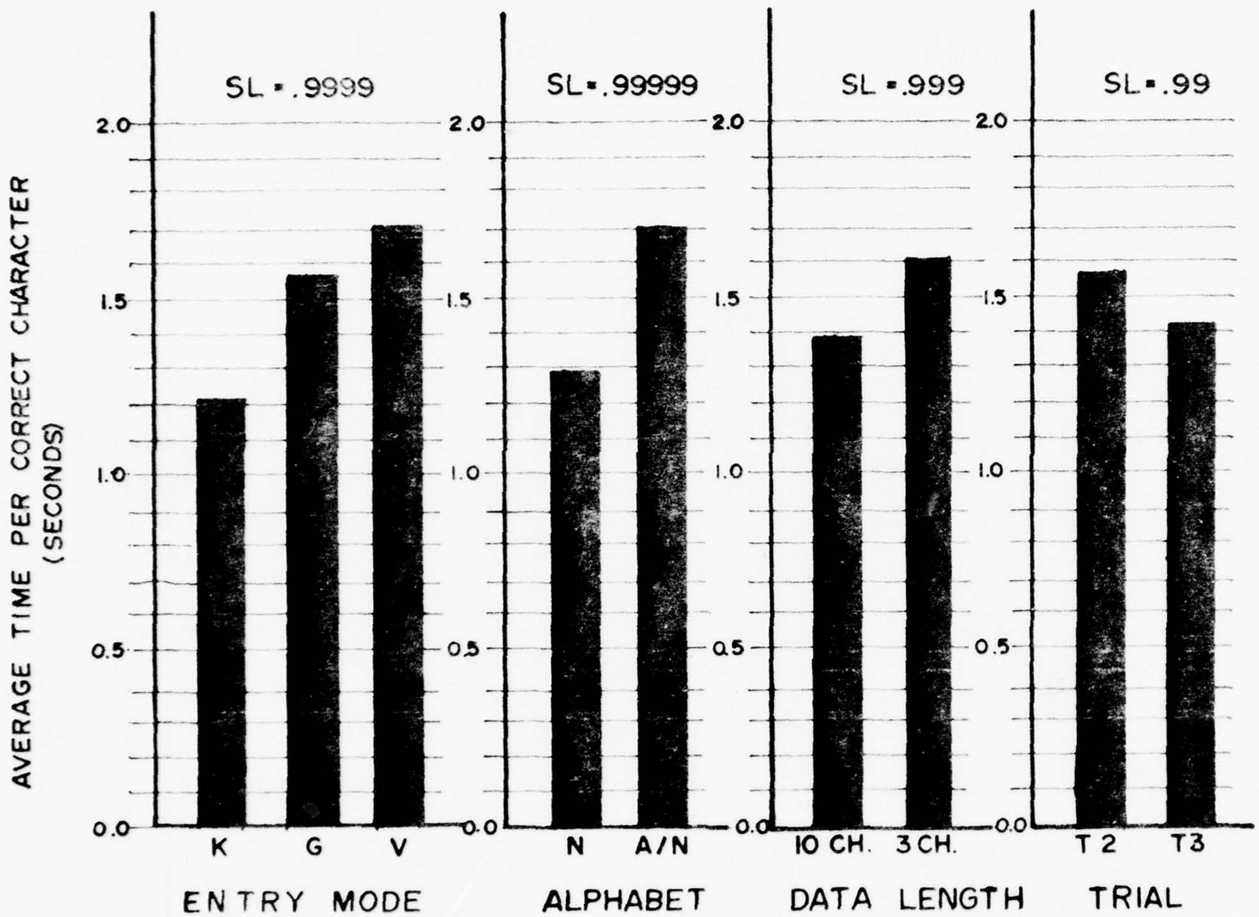


Fig. 5 HSDE Individual Factor Entry Time Comparisons

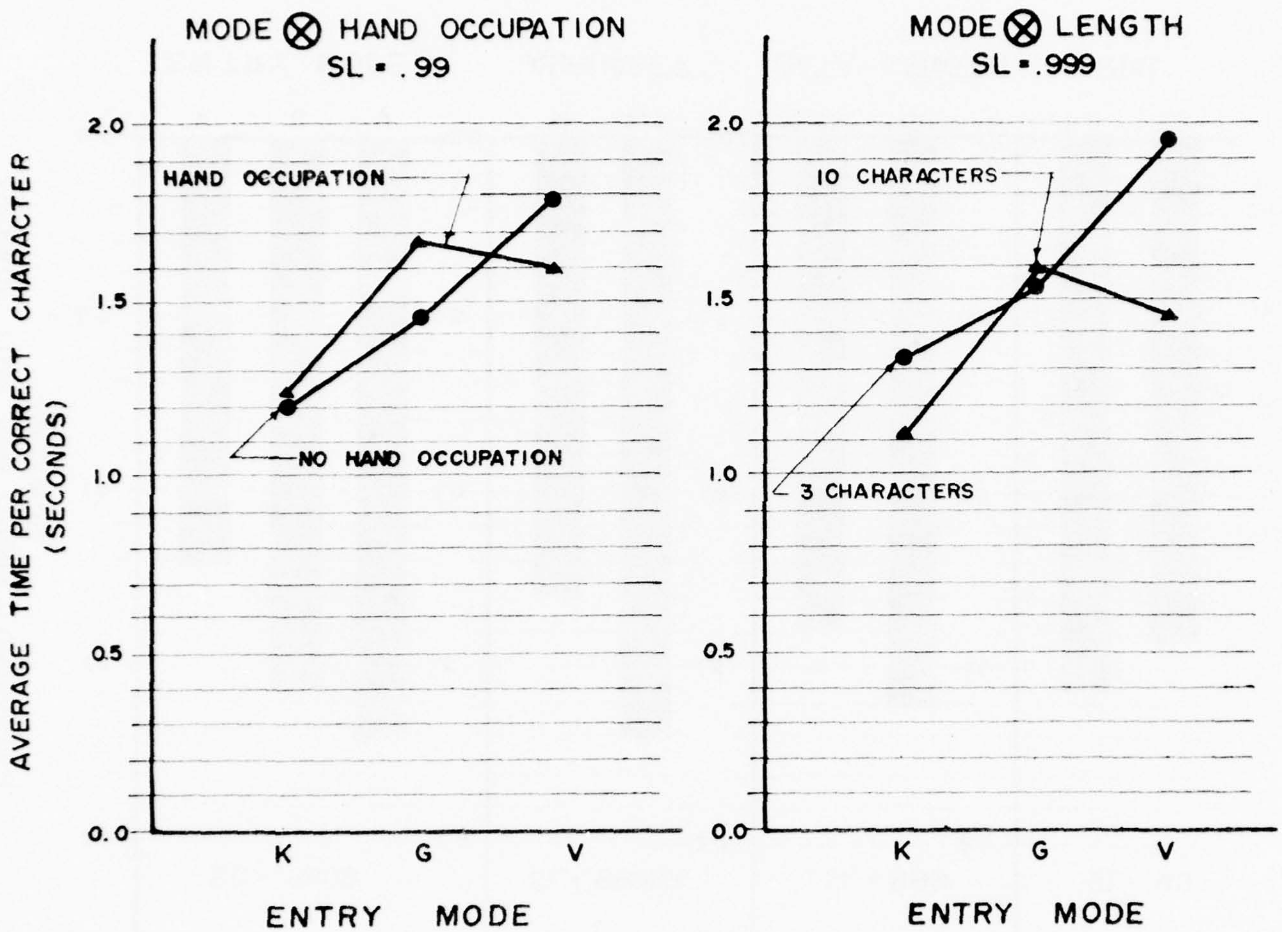


Fig. 6 Entry Time Interactions Between Entry Mode and Two Other Factors

pation. Significant interactions between two factors always show up in a graph in which one factor is the ordinate and the other is a parameter by lack of parallelism or crossing of the curves for the parameters.

The interaction between mode and length is significant because 10 character strings had substantially lower entry times than 3 character strings for voice entry, moderately lower for keyboard, but had higher times for Graf Pen.

The interaction of entry mode and hand occupation shows that hand occupation had no effect on keyboard entry, reduced the speed of Graf Pen entry and increased the speed of voice entry. The hand occupation requirement in this experiment was not significant, overall, primarily because it was such a simple task that it could be performed in a fraction of a second. Whatever negative effect it did have, however, would be expected to be greatest with Graf Pen for which it was necessary either for the subject to carry the data entry stylus back and forth between the data entry tablet and the pushbuttons or to lay down the stylus while pushing the buttons. On the other extreme, its negative effect would be expected to be least with voice input with which hands were not used at all for data entry. In fact, voice input proceeded faster on the average with hand occupation than without.

Figure 7 is a plot of entry times for the three-way interaction between hand occupation, alphabet, and data length. This plot illustrates a kind of threshold effect in the nearly trivial hand occupation requirement. For the relatively difficult task of entering alphanumeric data, the hand occupation requirement was so simple that it would not be expected to increase significantly the entry time per character. (In fact, the average entry time was decreased slightly with hand occupation.) For entry of 10 character numeric strings likewise, pushing the hand occupation buttons once per string was such a minor part of the total task that it would not be expected to increase significantly the entry time per character. For entry of 3 character numeric strings, however, pushing the buttons once per string increased the overhead per character by enough to significantly increase entry time per character.

In this interaction, it is difficult to explain how hand occupation could actually decrease entry time for 3 of the 4 conditions. It is most likely the result of uncontrolled intersubject variations, but it is also possible that a simple hand occupation task such as the one used in this experiment could improve data entry performance by making the task more rhythmic.

2. Error Rates After Correction

The operational errors, or errors after correction and verification are analyzed with respect to two closely related measures; percent correct characters and percent correct character strings. Generally, the significance levels for these measures are lower than for the time per correct character measurements. This reflects the higher degree of uncontrolled variations in these measures.

Table 12 summarizes the analysis of variance for Percent Correct

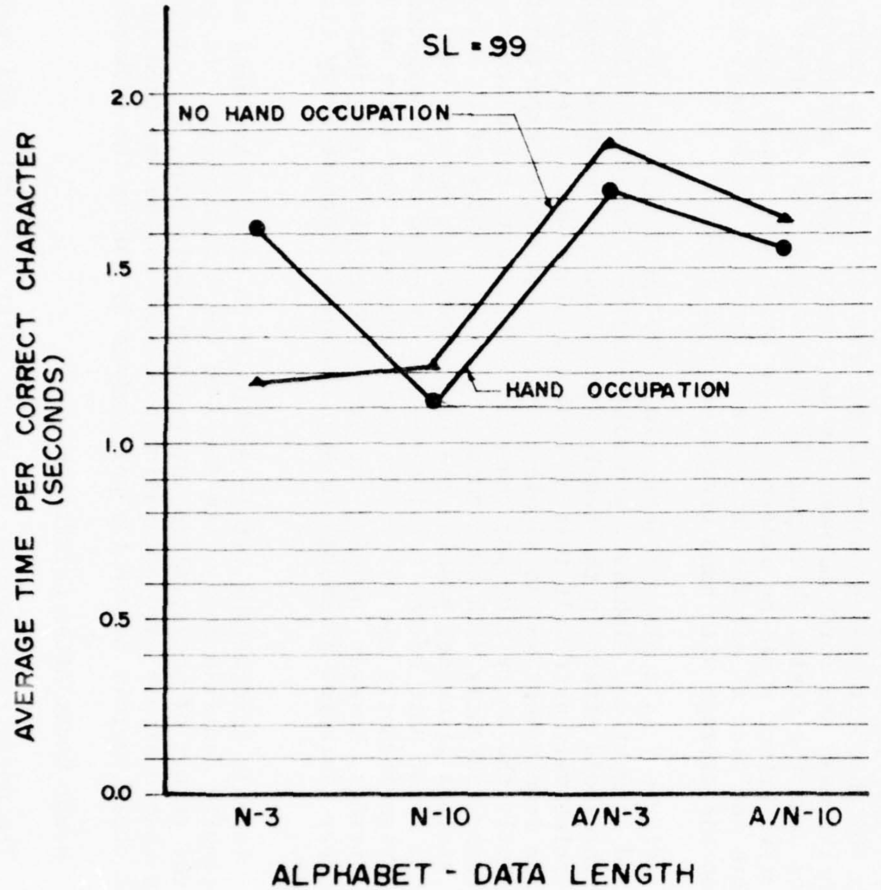


Fig. 7 Entry Time Interaction Between Hand Occupation, Alphabet and Data Length

TABLE 12

ANALYSIS OF VARIANCE OF
PERCENT CORRECT CHARACTERS

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	FEEDBACK (F)	1	4.698	4.698	-	-
2.	HAND OCCUPATION (H)	1	1.636	1.636	-	-
3.	FxH	1	1.499	1.499	-	-
4.	LENGTH (L)	1	2.447	2.447	-	-
5.	FxL	1	6.254	6.254	-	-
6.	HxL	1	1.111	1.111	-	-
7.	ALPAHBET (A)	1	20.318	20.318	7.58	.99
8.	FxA	1	13.451	13.451	5.02	.95
9.	HxA	1	0.441	0.441	-	-
10.	LxA	1	17.424	17.424	6.50	.95
11.	ENTRY MODE (M)	2	2.080	1.040	-	-
12.	FxM	2	8.801	4.400	-	-
13.	HxM	2	8.620	4.310	-	-
14.	LxM	2	4.079	2.039	-	-
15.	AxM	2	17.066	8.533	3.18	.95
16.	TRIAL (T)	2	22.897	11.448	4.27	.95
17.	FxT	2	0.796	0.398	-	-
18.	HxT	2	1.531	0.766	-	-
19.	LxT	2	0.452	0.226	-	-
20.	AxT	2	0.240	0.120	-	-
21.	MxT	4	3.165	0.791	-	-
22.	HxLxA	1	28.471	28.471	10.62	.995
	ALL INTERACTIONS BETWEEN 3, 4, 5 AND 6 FACTORS = ERROR	109	292.16	2.68		
	TOTAL	143	431.184			

GRAND MEAN = 99.036

Characters. For this measure the most significant result is a three-way interaction between hand occupation, length and alphabet. Alphabet and trial are the only two individual factors which are significant (at the .99 and .95 levels respectively). The two-way interactions between alphabet and feedback, alphabet and length, and alphabet and mode are significant at the 0.95 level.

Table 13 summarizes the analysis of variance for Percent Correct Character Strings. The results are predictably similar to those for Percent Correct Characters. The most significant result is once again a three-way interaction between hand occupation, length, and alphabet. The individual factors of alphabet and length are significant at the 0.995 level, and trial is significant at the 0.95 level. Length of data strings is more important to percent correct character strings than it is to percent correct character, since for a given character error rate, the string error rate will tend to be proportional to the string length. Finally, interactions between alphabet and hand occupation and alphabet and entry mode are both significant at the 0.95 level.

It is particularly noteworthy that neither feedback, hand occupation nor entry mode are significant in either of these measures of operational error rate. Both feedback and entry mode are significant, however, in interaction with alphabet.

Figure 8 compares the significant single factors for percent wrong characters after correction. Figure 9 compares the significant single factors for percent wrong character strings after correction. The overall string error rate was about five times as high as the overall character error rate. Both figures show, however, that the error rate after correction was more than twice as great for alphanumeric data as for numeric-only data, and that the effect of training was greater from T1 to T2 than from T2 to T3. Finally, the string error rate was about two and one-half times as great for 10 character strings as for 3 character strings.

The primary reason that the error rate after correction was higher for alphanumeric data than for numeric-only data was that with alphanumeric input there were numerous confusions between the characters S and 5 and 1 and I when reading the Burroughs Self-Scan. Since these were reading errors, they were not generally corrected before verification.

Figure 10 plots the interaction between entry mode and alphabet both for character errors and string errors. The plots are similar and both indicate that for alphanumeric data, voice input had approximately one-half the operational error rate of either keyboard or Graf Pen. For numeric-only data, the situation was nearly reversed. Keyboard had an extremely low error rate. Graf Pen had a somewhat higher error rate, and voice had a much higher rate.

Figure 11 plots the interaction of feedback and alphabet both for character errors and string errors. For numeric data, voice response had no significant effect on operational error rate. For alphanumeric data, however, voice response feedback more than doubled the operational error rate.

TABLE 13

ANALYSIS OF VARIANCE OF
PERCENT CORRECT CHARACTER STRINGS

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	FEEDBACK (F)	1	64.0	64.0	-	-
2.	HAND OCCUPATION (H)	1	36.0	36.0	-	-
3.	FxH	1	32.1	32.1	-	-
4.	LENGTH (L)	1	498.8	498.8	8.74	.995
5.	FxL	1	21.8	21.8	-	-
6.	HxL	1	16.0	16.0	-	-
7.	ALPHABET (A)	1	529.0	529.0	9.28	.995
8.	FxA	1	196.0	196.0	3.43	.90
9.	HxA	1	256.0	256.0	4.49	.95
10.	LxA	1	9.0	9.0	-	-
11.	ENTRY MODE (M)	2	61.6	30.8	-	-
12.	FxM	2	146.0	73.0	-	-
13.	HxM	2	194.7	97.3	-	-
14.	LxM	2	194.9	97.4	-	-
15.	AxM	2	482.7	241.3	4.22	.95
16.	TRIAL (T)	2	406.2	203.1	3.56	.95
17.	FxT	2	32.7	16.3	-	-
18.	HxT	2	32.0	16.0	-	-
19.	LxT	2	86.2	43.1	-	-
20.	AxT	2	72.0	36.0	-	-
21.	MxT	4	178.8	44.7	-	-
22.	HxLxA	1	747.1	747.1	13.11	.999
	ALL INTERACTIONS BETWEEN 3, 4, 5 AND 6 FACTORS = ERROR	109	6216.0	57.0		
	TOTAL	143	9762.5			

GRAND MEAN = 95.194

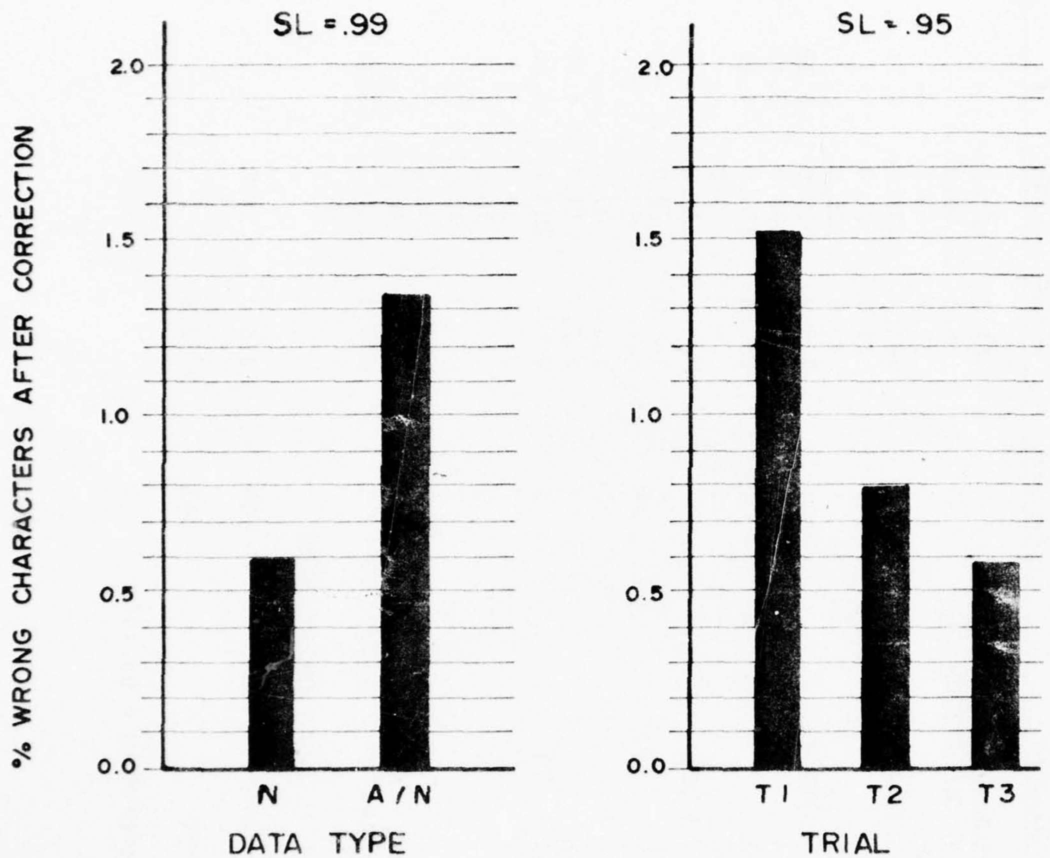


Fig. 8 Individual Factor Comparisons of Errors After Correction

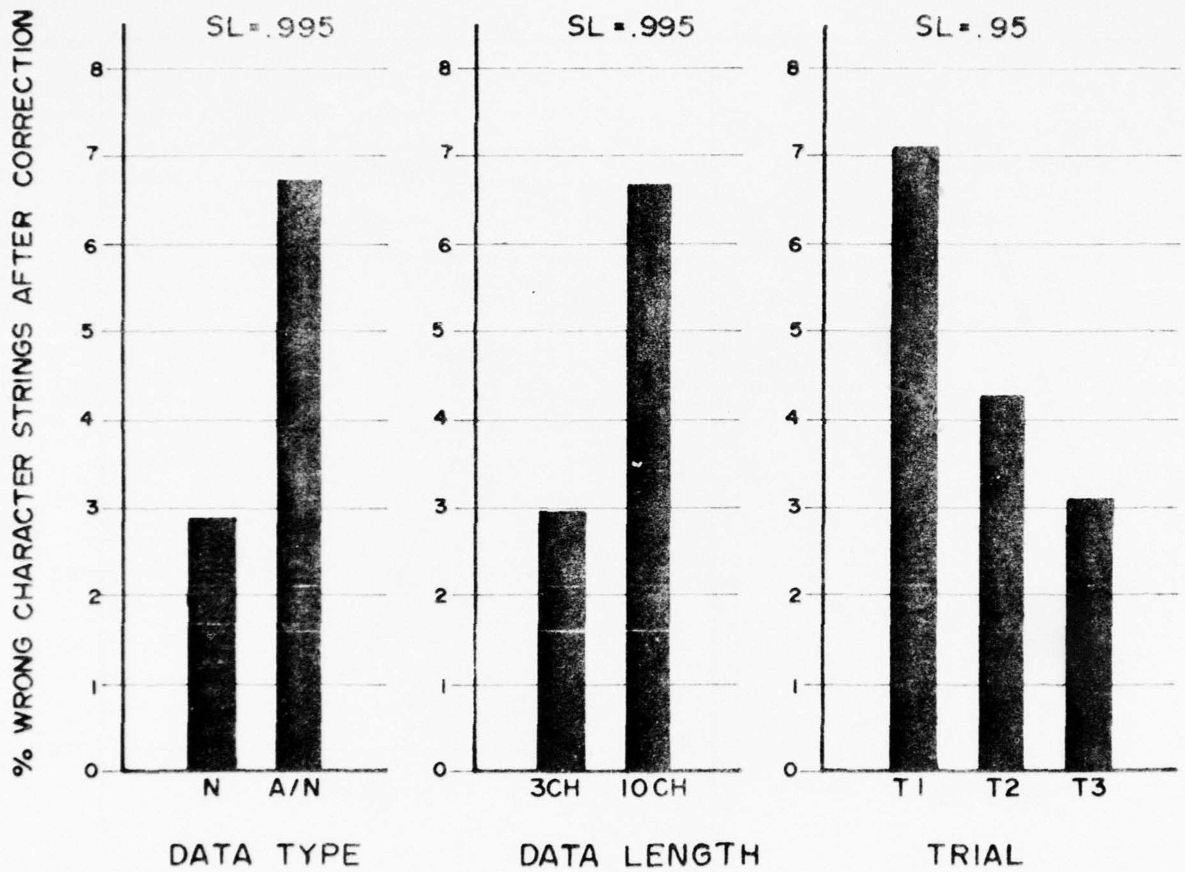


Fig. 9 Individual Factor Comparisons of String Errors After Correction

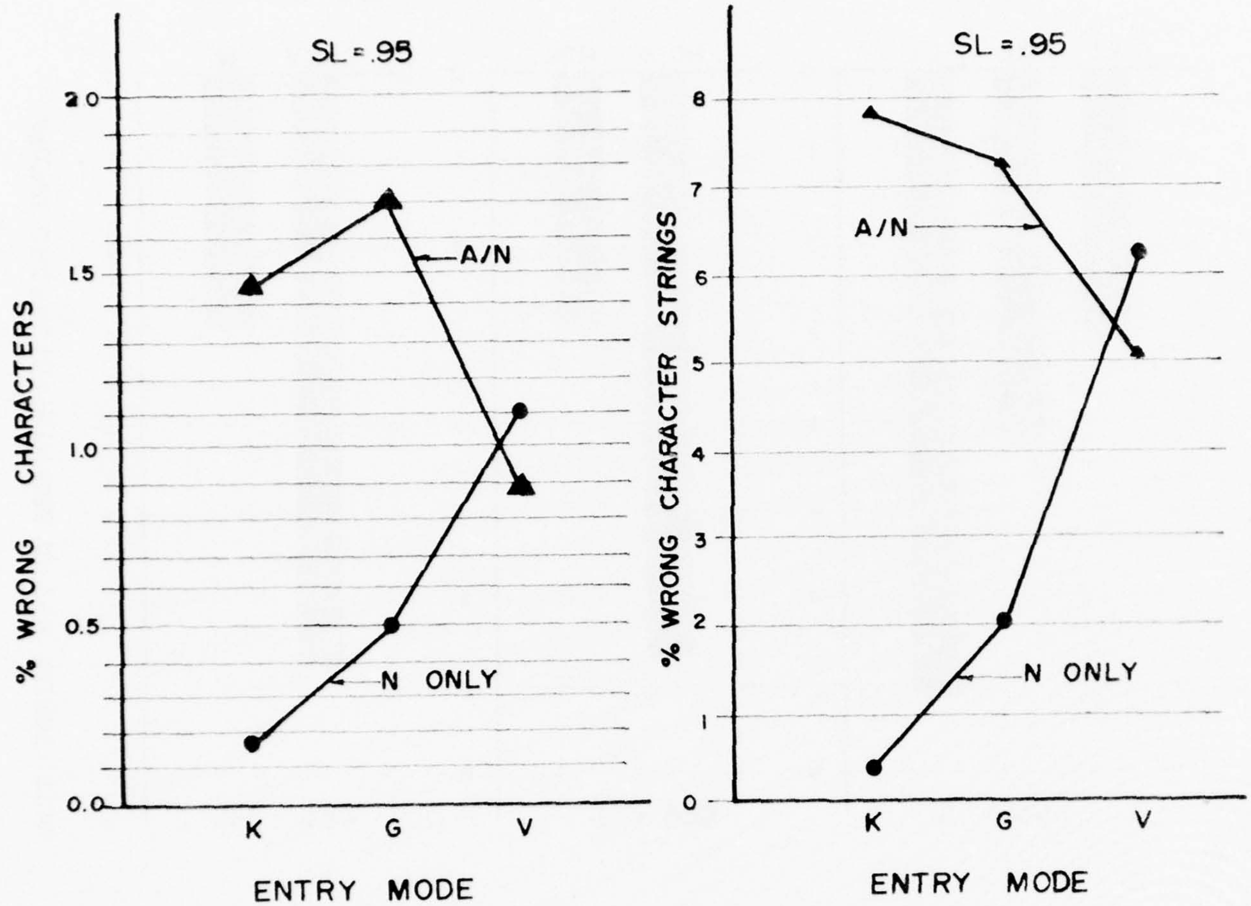


Fig. 10 Interactions Between Entry Mode and Alphabet for Character and String Errors

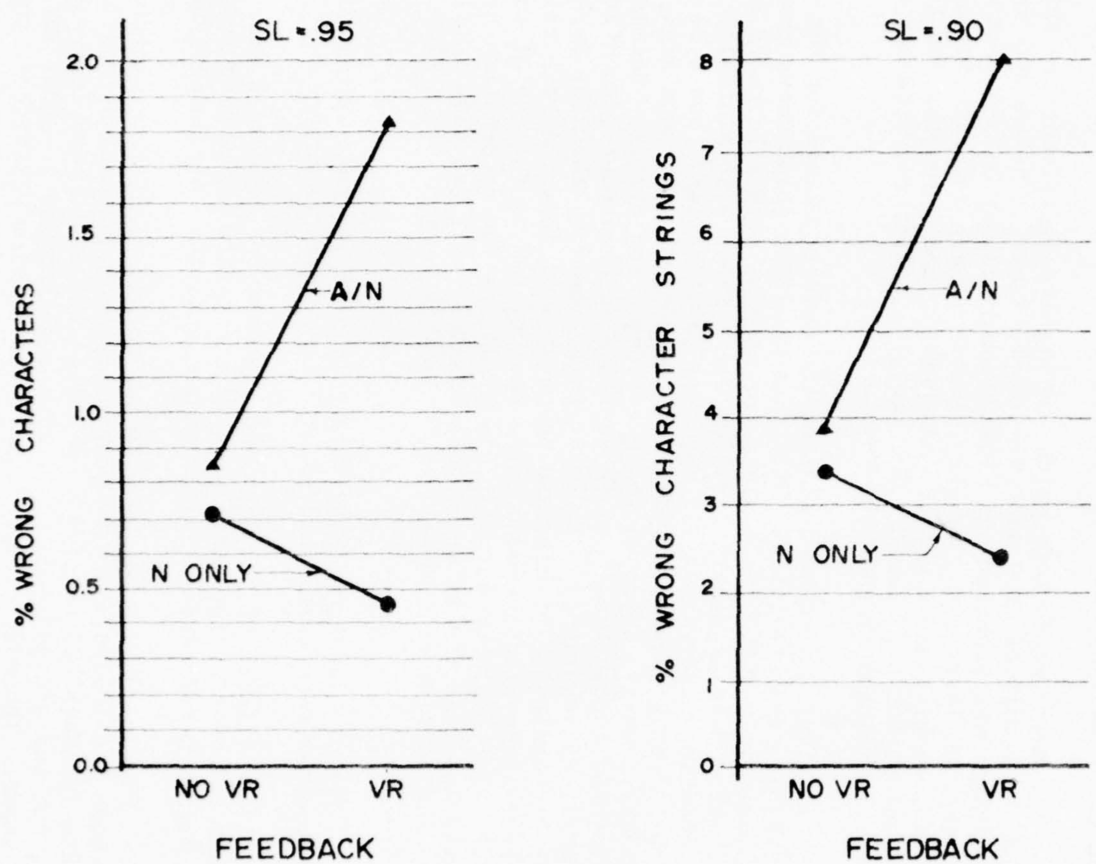


Fig. 11 Interaction Between Feedback Mode and Alphabet for Character and String Errors

3. Analysis of Basic Recognition Error Rate

Table 14 summarizes the analysis of variance for percent wrong characters before correction. This is a measure of the basic error rate of the data entry system and is only partly reflected in the operational error rate since many of these errors were corrected before final verification. This error measure is important since it gives an indication of the difficulty encountered by the subjects when using different entry devices.

The only two factors which are highly significant are entry mode and trial at levels of .999 and .995 respectively. The two-way interactions between length and alphabet and between hand occupation and mode are significant at the .95 level. And finally, two three-way interactions between hand occupation, length, and alphabet, and between mode, length, and alphabet are significant at the 0.95 level. The individual factors, feedback and alphabet are both significant at the 0.90 level.

In Figure 12 we have plotted the average values for the four factors which exhibit some statistical significance. The graph shows that voice entry had approximately twice the basic error rate of either keyboard or Graf Pen. A basic error rate of nearly 2.7% for voice input with minimally trained subjects operating under stress is not hard to understand, but 1.2% and 1.5% error rates for keyboard and Graf Pen respectively may seem high, since neither of these devices are supposed to make recognition errors. Because the errors were measured automatically in these tests, a breakdown into different types of error is not available. It is certain, however, that the keyboard did not produce "recognition" errors, and below we will show that the relatively high human error rate with keyboard entry was related to the use of voice response feedback. The Graf Pen actually did produce a few recognition errors early in the tests due to a faulty microphone assembly. It was also inclined to encourage "keying" errors as a result of the offset between the stylus tip and the spark gap.

The plot of error rate versus trial shows that substantial reductions in basic error rate were achieved with increasing experience.

The effects of feedback and alphabet on error rate were less significant but indicate an increase in basic error rate with voice response feedback and a higher error rate for alphanumeric data than for numeric-only data.

Figure 13 is a plot of the two weakly significant interactions between entry mode and feedback and between entry mode and hand occupation. Voice response feedback had no effect at all on the error rate for voice or Graf Pen input, but was accompanied by a very large increase in error rate with keyboard input. The voice response unit was so slow relative to the keyboard entry rate that almost all subjects went ahead of the feedback and tried to ignore it. It is possible that hearing the names of previously entered characters spoken while trying to enter a new character may have caused the higher basic error rate.

The interaction of hand occupation and mode indicates that hand occupation reduced the error rate for both voice and Graf Pen data entry, but

TABLE 14

ANALYSIS OF VARIANCE OF
PERCENT WRONG CHARACTERS BEFORE CORRECTION

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	FEEDBACK	1	11.816	11.816	3.69	.90
2.	HAND OCCUPATION (H)	1	8.023	8.023	-	-
3.	FxH	1	0.341	0.341	-	-
4.	LENGTH (L)	1	1.598	1.598	-	-
5.	FxL	1	8.995	8.995	2.81	.90
6.	HxL	1	0.001	0.001	-	-
7.	ALPAHBET (A)	1	10.096	10.096	3.15	.90
8.	FxA	1	1.762	1.762	-	-
9.	HxA	1	7.604	7.604	-	-
10.	LxA	1	16.477	16.477	5.15	.95
11.	ENTRY MODE (M)	2	60.882	30.441	9.51	.999
12.	FxM	2	16.725	8.362	2.61	.90
13.	HxM	2	23.797	11.899	3.72	.95
14.	LxM	2	1.166	0.583	-	-
15.	AxM	2	6.511	3.255	-	-
16.	TRIAL (T)	2	40.037	20.018	6.25	.995
17.	FxT	2	3.730	1.865	-	-
18.	HxT	2	0.974	0.487	-	-
19.	LxT	2	7.156	3.578	-	-
20.	AxT	2	2.229	1.115	-	-
21.	MxT	4	0.983	0.246	-	-
22.	HxLxA	1	15.009	15.009	4.69	.95
23.	LxAxM	2	24.699	12.349	3.86	.95
	ALL INTERACTIONS BETWEEN 3, 4, 5 AND 6 FACTORS = ERROR	109	348.87	3.20		
	TOTAL	143	597.768			

GRAND MEAN = 1.765

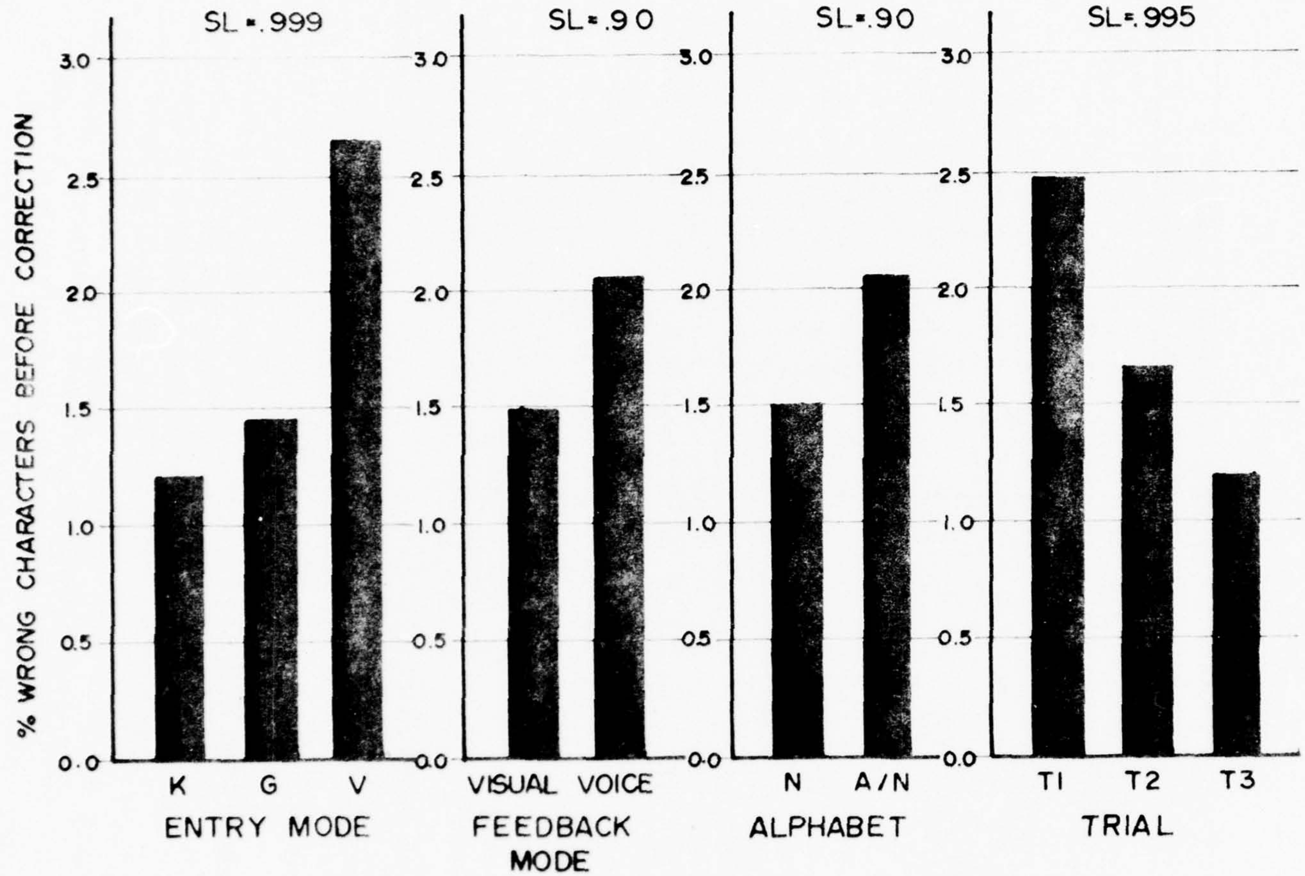


Fig. 12 Individual Factor Comparisons of Errors Before Correction

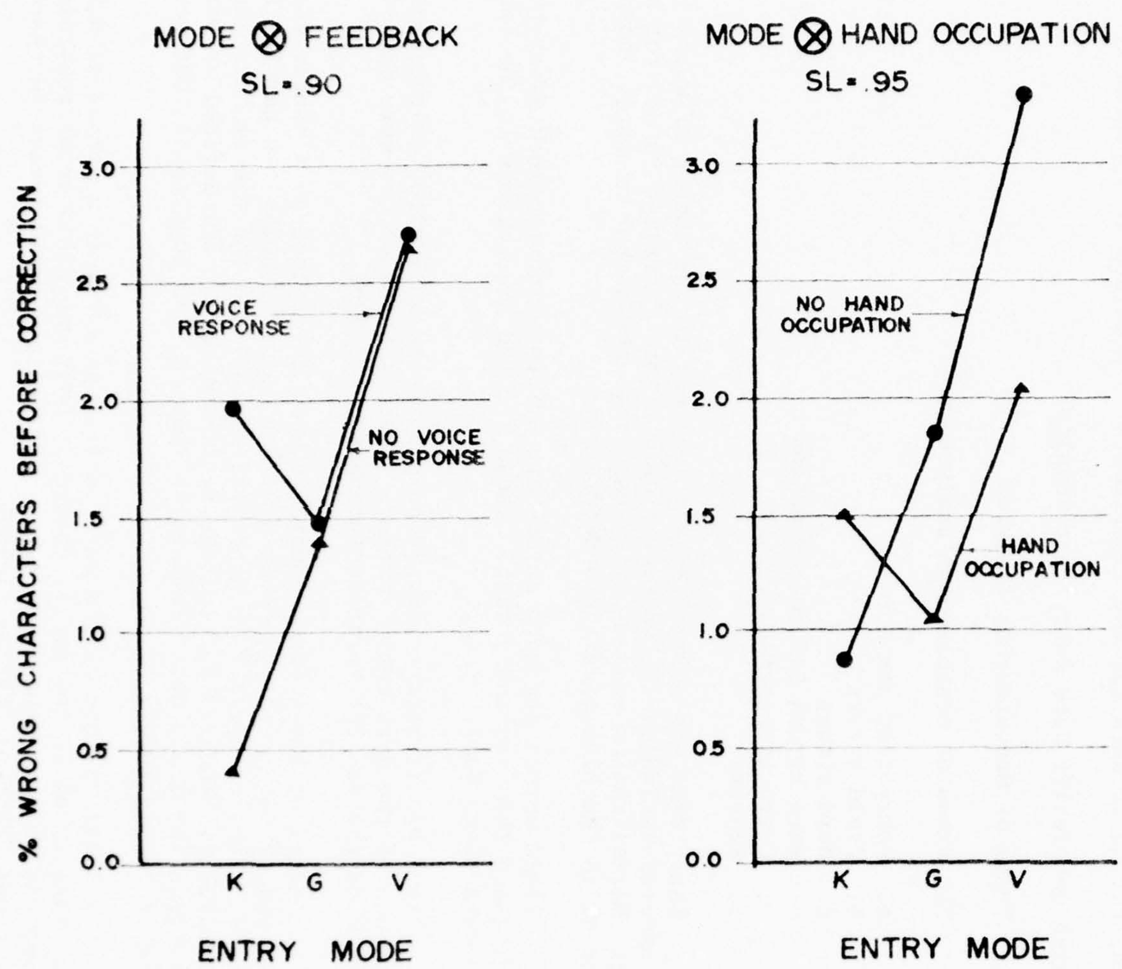


Fig. 13 Interaction Between Entry Mode and Two Other Factors for Errors Before Correction

increased it for keyboard entry. We have no explanation for this result and are inclined to attribute it to random inter-subject variations.

There are a number of other interactions which show some significance with respect to percent wrong characters before correction, but which we have not plotted. The reason for omitting them is that their significance levels are not extremely high and when plotted, they appear either not very interesting or not plausible. Should the reader wish to examine these interactions on his own, the basic data for the experiments is provided in Appendix A.

C. High Complexity Data Entry Test Results

1. Types of Measurements Analyzed

The types of measurements analyzed are:

a. Entry time per word
b. Field errors
c. Word errors
d. Word errors before correction
e. Correction system errors
f. Rejects

Entry time was normalized with respect to the number of words required to be entered including the end verification word (carriage return for keyboard). Normalization was not, with respect to the number of words, correctly entered as in the high speed data entry tests.

Field errors and word errors are errors which remained after the subject finished the data entry task. These are the true errors in the context of the data entry task.

Word errors before correction provide an indication of the basic error performance of the data entry systems. Most of these errors were detected and corrected by the subject before entering the messages.

In all of these measurements, a word is defined as a single entry in the context of either the voice of Graf Pen input systems. In particular, each data field name, each digit in numerical fields, and the end or "carriage return" character required for message verification are all defined as words. For keyboard input all data field names required two keystrokes, but were still defined as single words.

A field differs from a word in that an error in a string of numbers such as time or an ID number is counted as only one field error regardless of how many digits are actually in error. Percentage field errors furthermore, do not count "END" or "CR" as fields.

The three word and field error measurements have been further subdivided into:

a. Keying, recognition, and correction system errors
b. Reading and interpretation errors

Detection of errors and division of errors into subcategories was done manually by comparing a hardcopy of the subjects' responses to a set of known correct problem responses. Total error counts resulting from this procedure are well defined, but subdivision into categories is sometimes questionable.

In general, errors were specified as keying and recognition errors whenever the error seemed like a possible confusion response for voice input or a neighboring key or menu error for keyboard and Graf Pen and when the context of the error did not indicate that the particular character in error was simply part of a larger interpretation error or a simple reading error (such as confusion of 3 and 8). Errors were specified as correction system errors whenever they seemed to involve erroneous recognitions of the backspace and erase commands, or failures to respond to either of these commands, or when it was indicated by context that the error was a result of incorrect use of one of the correction commands.

Reading and interpretation errors included confusion of the order of data fields in the message, extraction of a data field from a neighboring message, and likely reading confusions such as between 3 and 8.

Rejects were provided for all three data entry devices whenever the data entry program detected illegal syntax. The entry system was highly structured, so that numerous rejects were obtained. For voice entry, rejects were also generated when the voice recognition system failed to recognize a word as one of the syntactically selected set of candidates. This could happen even though the correct word was spoken.

2. Entry Time Analysis

Table 15 summarizes the analysis of variance for the Time Per Word measurements in the high complexity data entry tests. Almost all of the variance is attributable to four factors and one interaction between two of these factors. The four factors are experience, hand occupation, entry mode and trial. Prompting was clearly not significant in these tests. The one highly significant interaction was between experience and entry mode.

The mean square error in this test is very low. This implies that virtually all of the variance in the test is attributable to the basic factors and two-way interactions between those factors. The significance levels in this test are higher than in the equivalent high speed data entry tests primarily because subject experience was made an explicit factor. Experience and trial are the two most significant factors in this test. If subject experience had been randomized as in the high speed data entry tests, its contribution to the variance would have appeared in the mean squared error term, and the F ratios of all of the other factors would have been reduced by more than one half.

Figure 14 graphs the high complexity entry times per word for the four factors which are statistically significant. Notice that in spite of the very high statistical significance, the numerical differences are not strikingly large. The entry mode comparison shows that Graf Pen and voice entry both required about 2.3 seconds per word, while keyboard required 2.94 seconds per

TABLE 15

ANALYSIS OF VARIANCE OF
HIGH COMPLEXITY DATA ENTRY TIME PER WORD

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERIENCE (E)	1	5.894	5.894	86.2	.99999
2.	HAND OCCUPATION (H)	1	3.957	3.957	57.9	.99999
3.	ExH	1	0.537	0.537	7.9	.99
4.	PROMPTING (P)	1	0.017	0.017	-	-
5.	ExP	1	0.005	0.005	-	-
6.	HxP	1	0.015	0.015	-	-
7.	ENTRY MODE (M)	2	6.231	3.115	45.5	.9999
8.	ExM	2	4.583	2.291	33.5	.9999
9.	HxM	2	0.643	0.321	4.7	.95
10.	PxM	2	1.031	0.516	7.5	.995
11.	TRIAL (T)	2	11.285	5.643	82.5	.99999
12.	ExT	2	0.023	0.011	-	-
13.	HxT	2	0.330	0.165	-	-
14.	PxT	2	0.054	0.027	-	-
15.	MxT	4	0.139	0.035	-	-
	ALL INTERACTIONS BETWEEN 3, 4, AND 5 FACTORS = ERROR	45	3.077	0.0684		
	TOTAL	71	37.822			

GRAND MEAN = 2.530

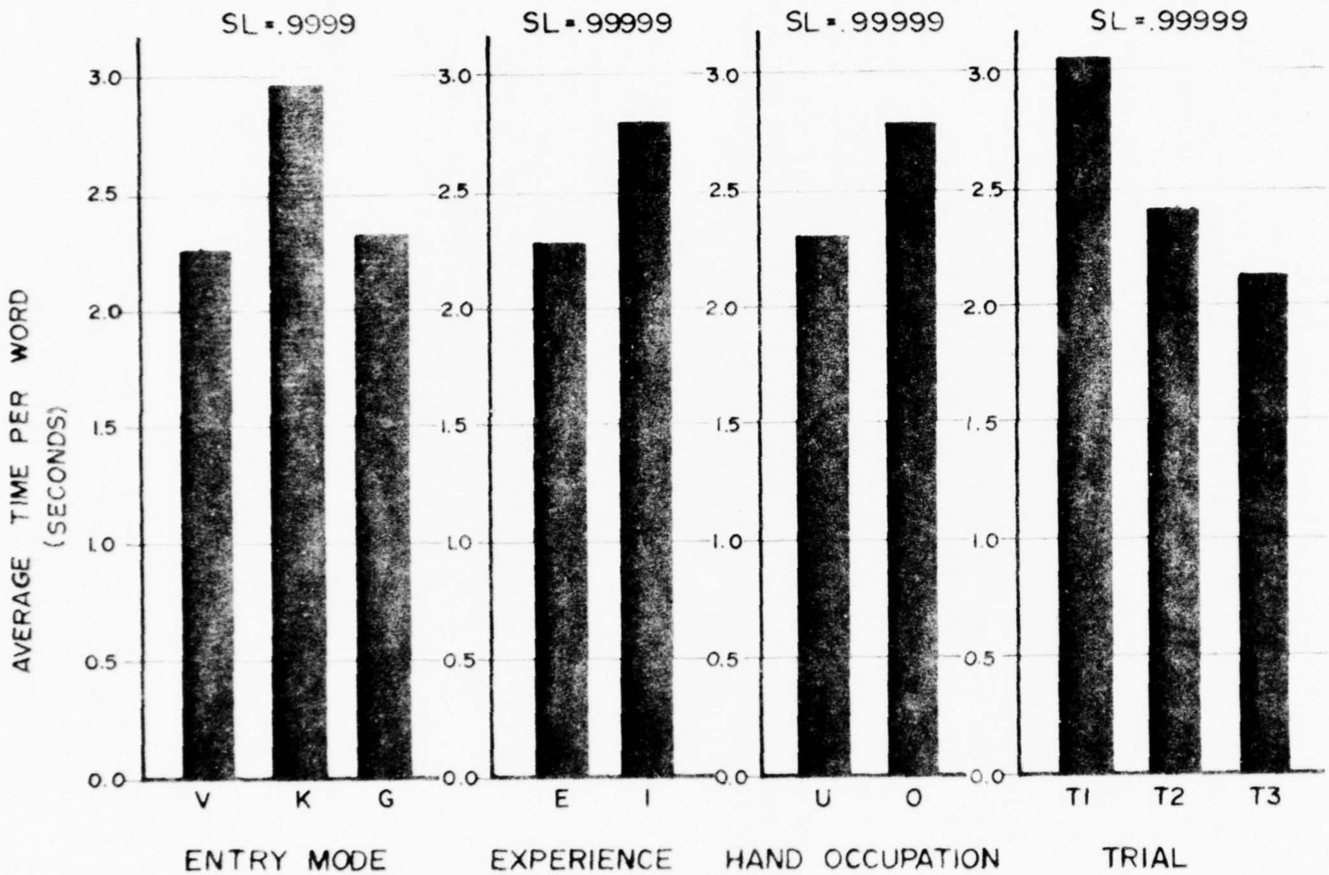


Fig. 14 HCDE Individual Factor Entry Time Comparisons

word. Thus, keyboard is about 30% slower. This difference is partly attributable to the requirement for entering two characters per non-numeric word on the keyboard with only one entry required for either voice or Graf Pen, and as we will show later, is a strong function of subject experience.

Experience, trial, and hand occupation all show the expected results. Experienced operators required about 20% less time than inexperienced operators, and the test time dropped by about 30% from Trial 1 to Trial 3. Hand occupation had a significant effect in this experiment because the 3.5 second occupation time was applied to each message. Since the average number of words per message was 7.5, the minimum increment would then be 0.47 seconds per word. Coincidentally, The average measured increment is 0.47 seconds per word, and as we will see below, the hand occupation increment was not as great for voice as it was for keyboard and Graf Pen.

Figure 15 displays the averages for the three interactions between entry mode and other variables. The most significant interaction is between mode and experience. Inexperience increased entry time by 56% for keyboard, by 14% for Graf Pen and only by 5% for voice. For all three devices, the inexperienced operators were totally inexperienced with the entry devices and the so-called experienced subjects were never experts. The tests indicate that it takes more than a little experience to make a big difference in entry time in this kind of test for either voice or Graf Pen entry, but the difference between no familiarity and some familiarity with the keyboard has a substantial effect on throughput, since, when used by totally inexperienced subjects, keyboard is quite a slow entry device.

The interaction of mode and hand occupation is significant at a much lower level than either of the individual factors. In the tests, hand occupation increased voice input time by 0.21 seconds per word, keyboard by 0.54 seconds per word, and Graf Pen by 0.66 seconds per word. As previously stated, the minimum time required for hand occupation is about 0.47 seconds per word unless that time can be absorbed into the entry time, as it could with voice input. This test has demonstrated an advantage to voice input when the hands are occupied, but once again, the hand occupation must be substantial, and even then, not all of the hand occupation time will be absorbed into the entry time for voice input.

An additional result was that Graf Pen was hurt slightly more by hand occupation than was keyboard. The greater slow down for the Graf Pen may have been caused because part of hhe Graf Pen mechanism had to be held in one of the operator's hands. Moving the Graf Pen stylus and cable back and forth between the entry tablet and the push buttons would logically be more time consuming than simply moving empty hands.

The interaction between entry mode and prompting can be summarized as follows: voice prompting increased entry time for voice by almost 17%, had no significant effect on keyboard entry, and decreased entry time for Graf Pen by about 9%. Voice prompting slowed down voice entry because the voice operators were almost always inclined to wait for the voice response unit to stop talking before they started talking. Furthermore, the capability of the voice prompting to free the subjects' eyes did not have its usual value, since voice input

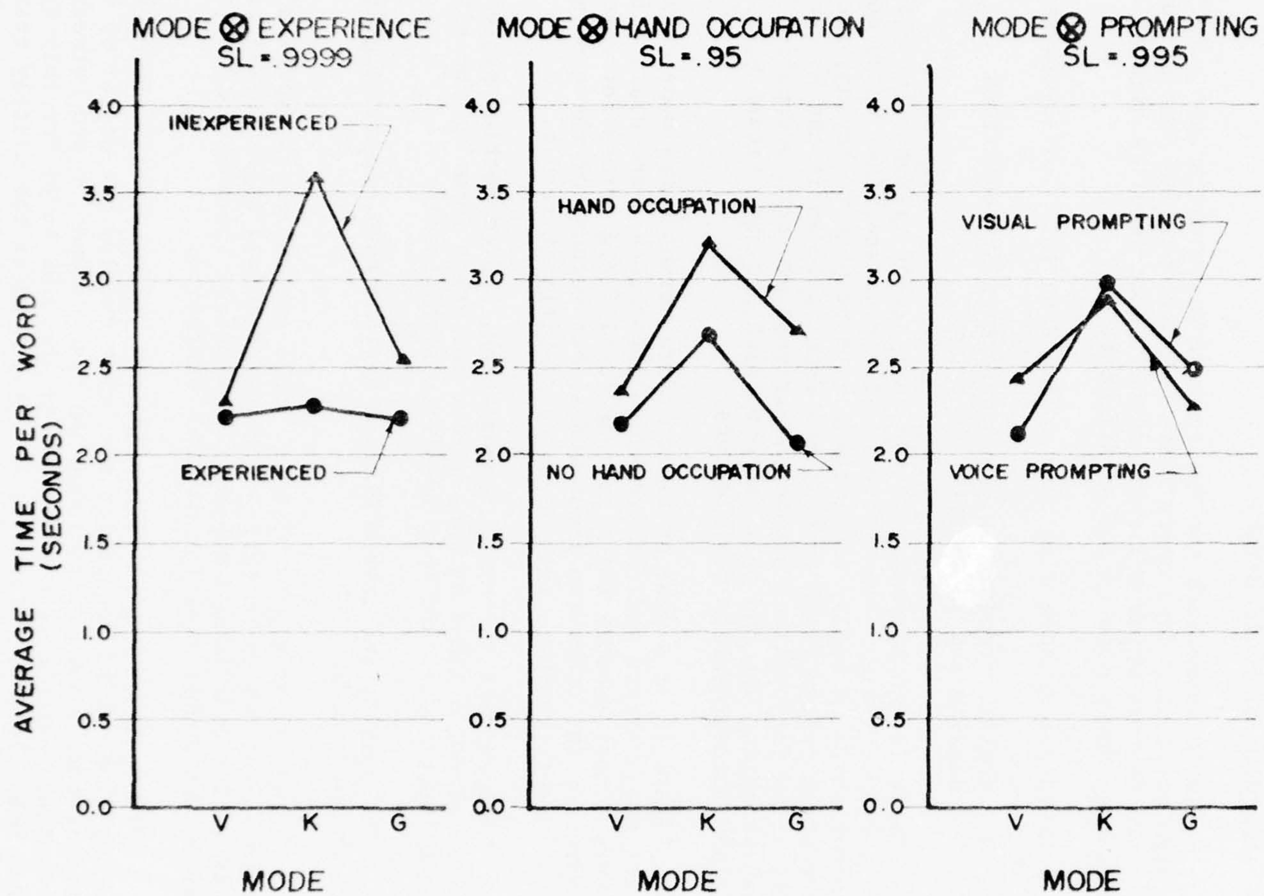


Fig. 15 Entry Time Interactions Between Entry Mode and Three Other Factors

also freed his eyes. For the Graf Pen, on the other hand, there was no hesitation to enter data while the VRU was talking; and since the eyes were very busy, voice prompting was quite helpful. It would have been even more helpful if it had had a more optimum vocabulary and if it had been modified for greater speaking speed.

3. Analysis of Field Errors

Table 16 summarizes the Analysis of Variance for the total field errors in the high complexity data entry tests. Experience and hand occupation are the only two factors which achieve a significance level as high as 0.90 for this measure. Entry mode is not significant.

Field errors have also been broken down into two subclasses:

a. Keying, recognition, and correction system errors
b. Reading and interpretation errors

Table 17 and 18 summarize the analysis of variance for class a) and class b), respectively. From Table 17 it can be seen that entry mode is the only factor which is really significant with respect to keying, recognition, and correction system field errors; and from Table 18 it can be seen that hand occupation and interactions between experience and hand occupation and experience and prompting are the only factors which are significant with respect to reading and interpretation field errors.

Figure 16 is a plot of field errors versus entry mode. This plot shows the total errors and the breakdown into class a and class b field errors. There were almost exactly twice as many total field errors for voice input as there were for the other two input modes, but because of the generally high variance in this measure, this is not a significant result.

With respect to keying, recognition, and correction systems errors, there were almost 5 times as many errors for voice input as for either of the other two input modes; and this result is significant.

The reading and interpretation field errors are not significant with respect to entry mode.

Figure 17 is a plot of total field errors versus experience and hand occupation. The number of errors for inexperienced subjects was almost twice as great as for experienced subjects. Likewise, the number of errors for hand occupation was double that for no hand occupation.

Figure 18 is a plot of reading and interpretation field errors versus hand occupation. Hand occupation more than tripled the number of these errors. The interactions between experience and hand occupation and between experience and prompting, which were significant at the 0.95 level for this type of error, have not been plotted. They will be covered under the closely related analysis of word errors in Section III-C-4.

TABLE 16

ANALYSIS OF VARIANCE OF
TOTAL FIELD ERRORS

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERIENCE (E)	1	1.681	1.681	3.23	.90
2.	HAND OCCUPATION (H)	1	1.681	1.681	3.23	.90
3.	ExH	1	0.681	0.681	-	-
4.	PROMPTING (P)	1	0.125	0.125	-	-
5.	ExP	1	0.681	0.681	-	-
6.	HxP	1	0.014	0.014	-	-
7.	ENTRY MODE (M)	2	2.528	1.264	2.4	.75
8.	ExM	2	2.528	1.264	2.4	.75
9.	HxM	2	0.194	0.039	-	-
10.	PxM	2	1.083	0.542	-	-
11.	TRIAL (T)	2	0.028	0.014	-	-
12.	ExT	2	0.361	0.181	-	-
13.	HxT	2	2.528	1.264	2.4	.75
14.	PxT	2	0.750	0.375	-	-
15.	MxT	4	3.556	0.889	-	-
	ALL INTERACTIONS BETWEEN 3, 4 AND 5 FACTORS = ERROR	45	23.567	0.524		
	TOTAL	71	41.986			

GRAND MEAN = 0.486

TABLE 17

ANALYSIS OF VARIANCE OF
KEYING RECOGNITION AND CORRECTION SYSTEM FIELD ERRORS

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERINCE (E)	1	1.125	1.125	3.39	.90
2.	HAND OCCUPATION (H)	1	0.125	0.125	-	-
3.	ExH	1	0.014	0.014	-	-
4.	PROMPTING (P)	1	0.014	0.014	-	-
5.	ExP	1	0.014	0.014	-	-
6.	HxP	1	0.681	0.681	-	-
7.	ENTRY MODE (M)	2	4.000	2.000	6.02	.99
8.	ExM	2	2.333	1.667	3.52	.95
9.	HxM	2	0.333	0.167	-	-
10.	PxM	2	0.444	0.222	-	-
11.	TRIAL (T)	2	0.750	0.375	-	-
12.	ExT	2	0.083	0.041	-	-
13.	HxT	2	0.250	0.125	-	-
14.	PxT	2	0.528	0.264	-	-
15.	MxT	4	1.250	0.313	-	-
	ALL INTERACTIONS BETWEEN 3, 4 AND 5 FACTORS = ERROR	45	14.94	0.332		
	TOTAL	71				

GRAND MEAN = 0.292

TABLE 18

ANALYSIS OF VARIANCE OF
READING AND INTERPRETATION FIELD ERRORS

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERIENCE (E)	1	0.056	0.056	-	-
2.	HAND OCCUPATION (H)	1	0.889	0.889	4.05	.95
3.	ExH	1	0.889	0.889	4.05	.95
4.	PROMPTING (P)	1	0.222	0.222	-	-
5.	ExP	1	0.889	0.889	4.05	.95
6.	HxP	1	0.500	0.500	-	-
7.	ENTRY MODE (M)	2	0.194	0.097	-	-
8.	ExM	2	0.028	0.014	-	-
9.	HxM	2	0.361	0.181	-	-
10.	PxM	2	0.194	0.097	-	-
11.	TRIAL (T)	2	0.528	0.264	-	-
12.	ExT	2	0.194	0.097	-	-
13.	HxT	2	1.361	0.681	3.10	.90
14.	PxT	2	0.028	0.014	-	-
15.	MxT	4	1.056	0.264	-	-
	ALL INTERACTIONS BETWEEN 3, 4 AND 5 FACTORS = ERROR	45	9.882	0.2196		
	TOTAL	71				

GRAND MEAN = 0.194

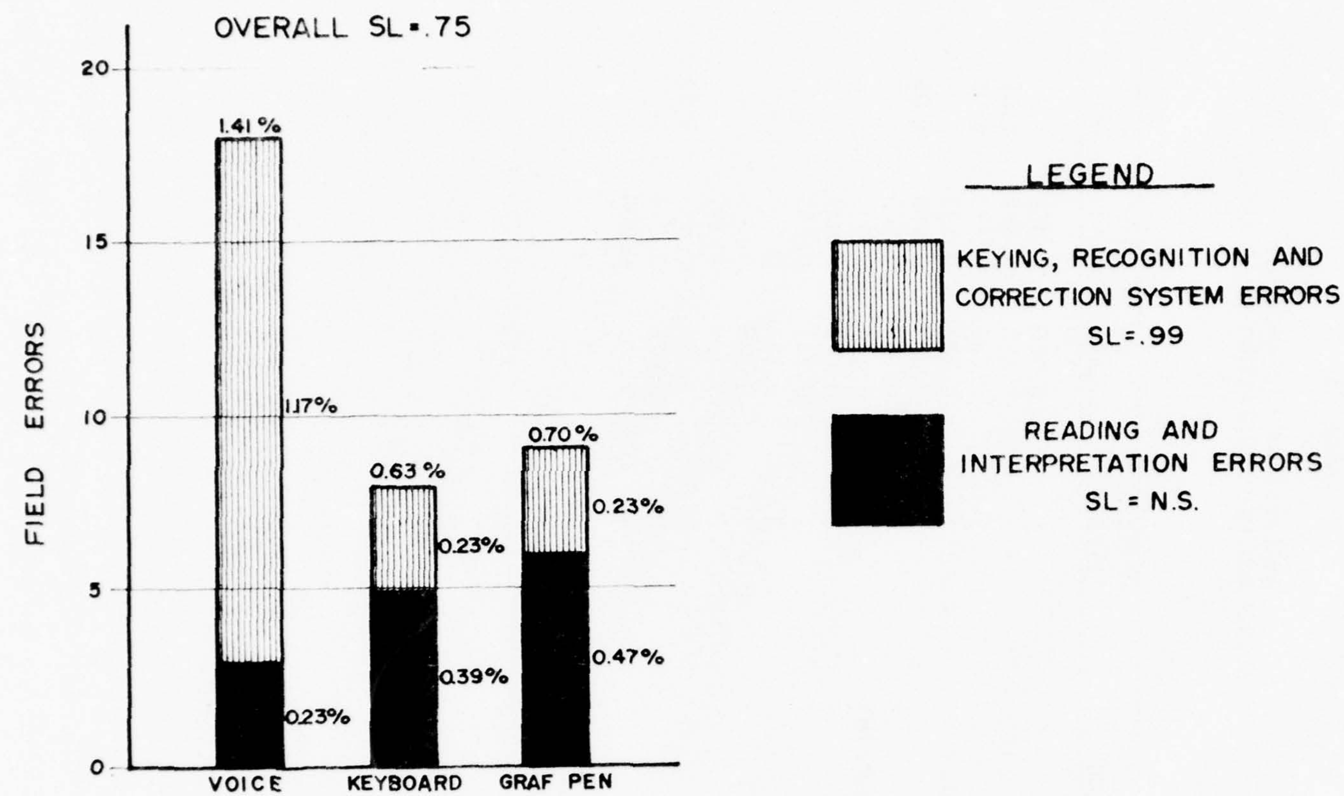


Fig. 16 Field Errors Versus Entry Mode

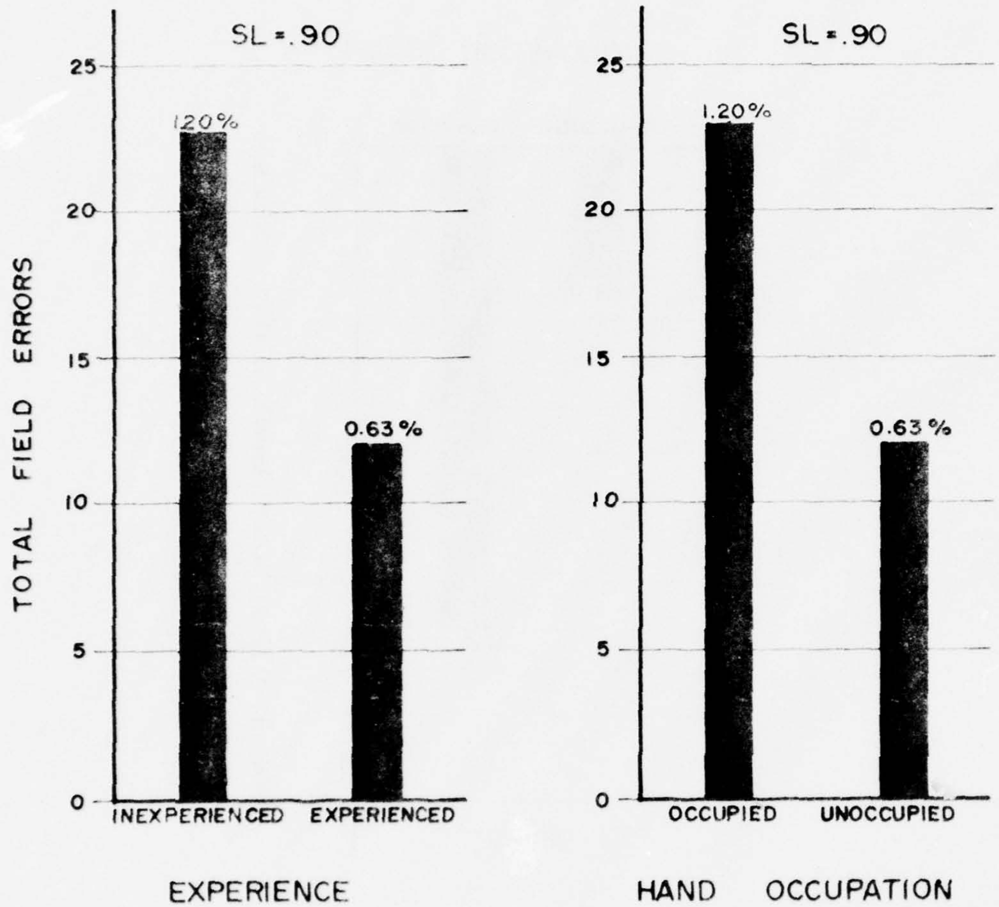


Fig. 17 Total Field Errors Versus Experience and Hand Occupation

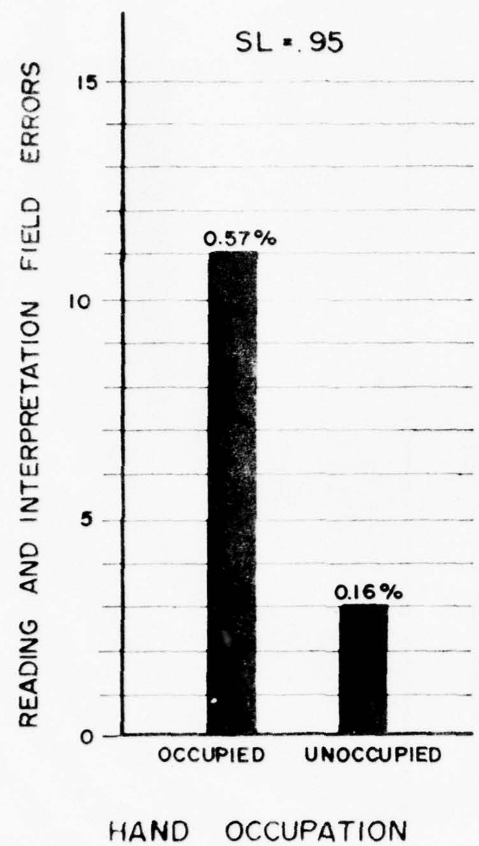


Fig. 18 Reading and Interpretation Field Errors Versus Hand Occupation

4. Analysis of Word Errors

Table 19 summarizes the analysis of variance for the total word errors in the high complexity data entry tests. Word errors have also been broken down into two subclasses:

a. Keying, recognition, and correction system errors
b. Reading and interpretation errors

Tables 20 and 21 summarize the analysis of variance for class a) and class b) errors respectively.

From Table 19 for total word errors, it can be seen that experience and hand occupation are the only significant single factors and that two additional two-factor interactions between experience and hand occupation and experience and prompting achieve some significance.

From Table 20 it can be seen that entry mode is the only factor which is really significant with respect to keying, recognition, and correction system word errors.

From Table 21 it can be seen that hand occupation and interactions between experience and hand occupation, experience and prompting, and prompting and hand occupation all have statistical significance with respect to reading and interpretation word errors.

Figure 19 is a graph of word errors versus entry mode. This graph shows the total errors and the breakdown into class a) and class b) word errors. The differences between total word errors are not significant, although voice input had slightly more of these errors than the other two input modes.

With respect to keying, recognition, and correction system errors, there were more than three times as many errors for voice input as for either of the other two input modes, and this result is significant.

Voice input produced about half as many reading and interpretation errors as the other two modes, but this difference was not great enough to achieve statistical significance.

Figure 20 is a plot of total word errors versus experience and hand occupation. The number of errors for inexperienced subjects or hand occupation was nearly two and one-half times as great as for experienced subjects and no hand occupation respectively.

Figure 21 a) and b) are plots of interactions between experience and hand occupation and between experience and prompting for total word errors. Hand occupation had no effect on total word errors with experienced subjects. With inexperienced subjects, however, hand occupation quadrupled the error rate as compared to no hand occupation. Figure 19 b) shows that voice prompting slightly increased the error rate with experienced subjects, but greatly decreased the rate with inexperienced subjects.

TABLE 19

ANALYSIS OF VARIANCE OF
TOTAL WORD ERRORS

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERIENCE (E)	1	7.347	7.347	5.4	.95
2.	HAND OCCUPATION (H)	1	7.347	7.347	5.4	.95
3.	ExH	1	7.347	7.347	5.4	.95
4.	PROMPTING (P)	1	3.125	3.125	-	-
5.	ExP	1	10.125	10.125	7.5	.99
6.	HxP	1	3.125	3.125	-	-
7.	ENTRY MODE (M)	2	0.444	0.222	-	-
8.	ExM	2	2.111	1.056	-	-
9.	HxM	2	0.778	0.389	-	-
10.	PxM	2	2.333	1.667	-	-
11.	TRIAL (T)	2	0.528	0.264	-	-
12.	ExT	2	4.694	2.347	-	-
13.	HxT	2	8.528	4.264	3.15	.90
14.	PxT	2	6.083	3.042	-	-
15.	MxT	4	8.389	2.097	-	-
	ALL INTERACTIONS BETWEEN 3, 4 AND 5 FACTORS = ERROR	45	60.66	1.35		
	TOTAL	71	132.99			

GRAND MEAN = 0.764

TABLE 20

ANALYSIS OF VARIANCE OF
KEYING, RECOGNITION AND CORRECTION SYSTEM WORD ERRORS

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERIENCE (E)	1	1.389	1.389	3.66	.90
2.	HAND OCCUPATION (H)	1	0.056	0.056	-	-
3.	ExH	1	0.056	0.056	-	-
4.	PROMPTING (P)	1	0.0	0.0	-	-
5.	ExP	1	0.0	0.0	-	-
6.	HxP	1	0.889	0.889	-	-
7.	ENTRY MODE (M)	2	3.694	1.847	4.87	.95
8.	ExM	2	2.028	1.014	-	-
9.	HxM	2	0.528	0.264	-	-
10.	PxM	2	0.583	0.292	-	-
11.	TRIAL (T)	2	0.528	0.264	-	-
12.	ExT	2	0.028	0.014	-	-
13.	HxT	2	0.194	0.097	-	-
14.	PxT	2	0.583	0.292	-	-
15.	MxT	4	1.556	0.389	-	-
	ALL INTERACTIONS BETWEEN 3, 4 AND 5 FACTORS = ERROR	45	17.17	0.38		
	TOTAL	71	29.278			

GRAND MEAN = 0.306

TABLE 21

ANALYSIS OF VARIANCE OF
READING AND INTERPRETATION WORD ERRORS

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERIENCE (E)	1	2.347	2.347	-	-
2.	HAND OCCUPATION (H)	1	6.125	6.125	5.52	.95
3.	ExH	1	8.681	8.681	7.82	.99
4.	PROMPTING (P)	1	3.125	3.125	-	-
5.	ExP	1	10.125	10.125	9.12	.995
6.	HxP	1	7.347	7.347	6.62	.95
7.	ENTRY MODE (M)	2	1.583	0.792	-	-
8.	ExM	2	0.528	0.264	-	-
9.	HxM	2	2.250	1.125	-	-
10.	PxM	2	0.583	0.292	-	-
11.	TRIAL (T)	2	0.333	0.167	-	-
12.	ExT	2	4.111	2.056	-	-
13.	HxT	2	6.333	3.167	-	-
14.	PxT	2	3.000	1.500	-	-
15.	MxT	4	5.583	1.396	-	-
	ALL INTERACTIONS BETWEEN 3, 4 AND 5 FACTORS = ERROR	45	49.82	1.11		
	TOTAL	71	111.875			

GRAND MEAN = 0.458

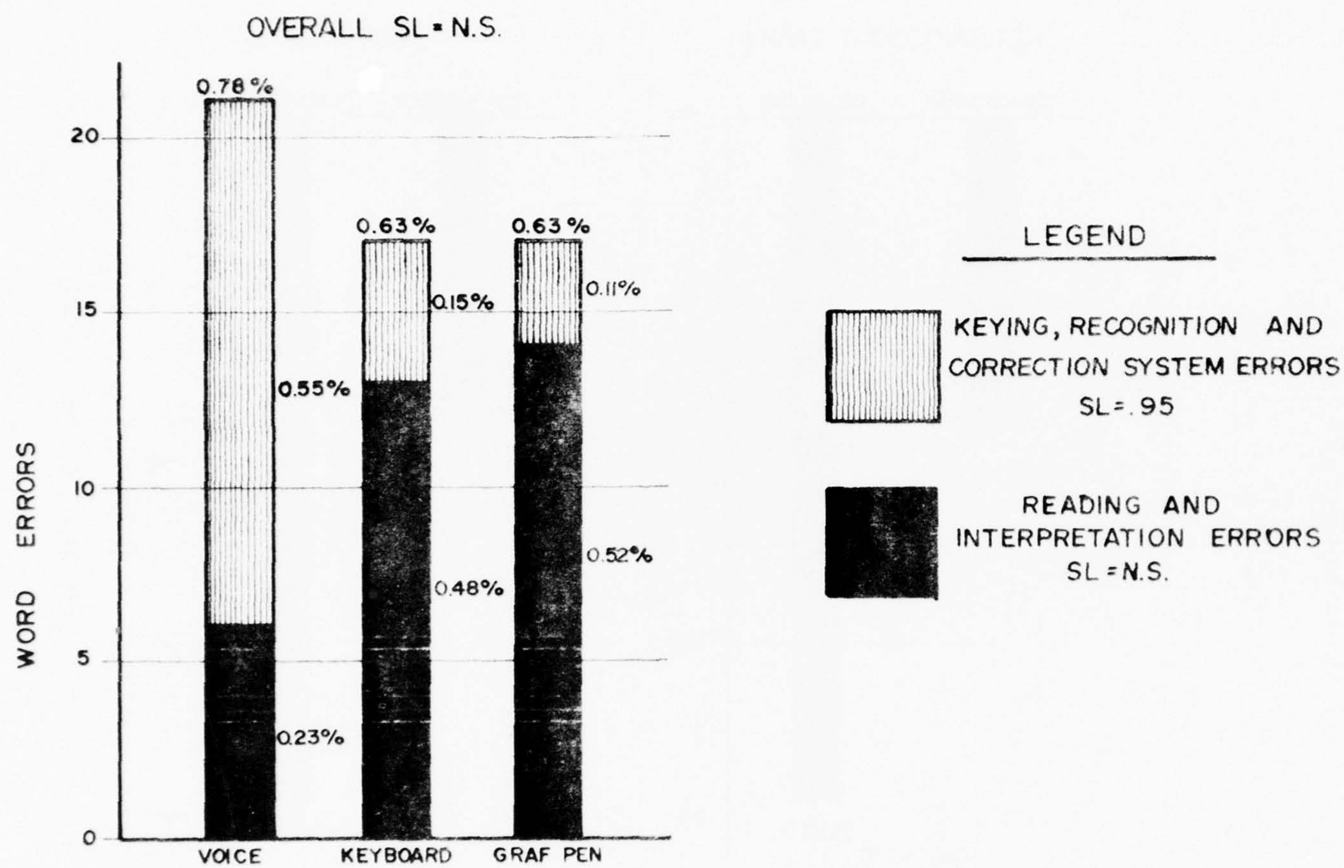


Fig. 19 Word Errors Versus Entry Mode

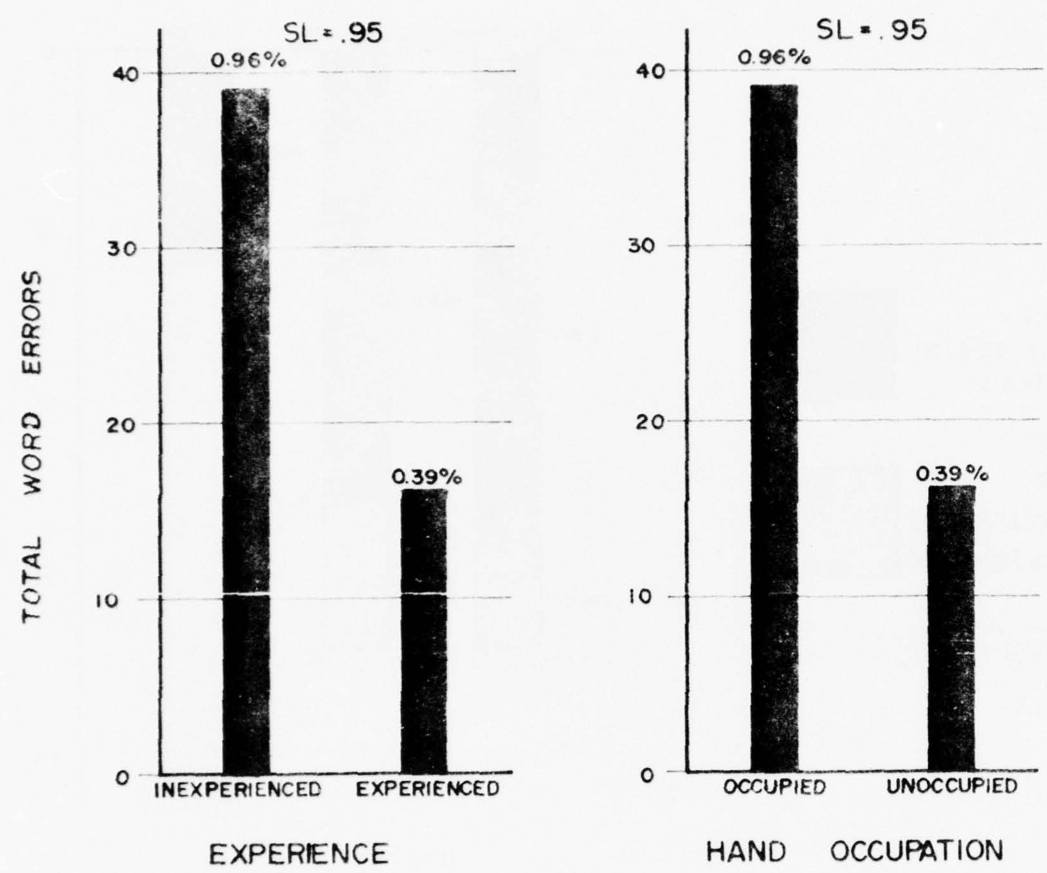


Fig. 20 Total Word Errors Versus Experience and Hand Occupation

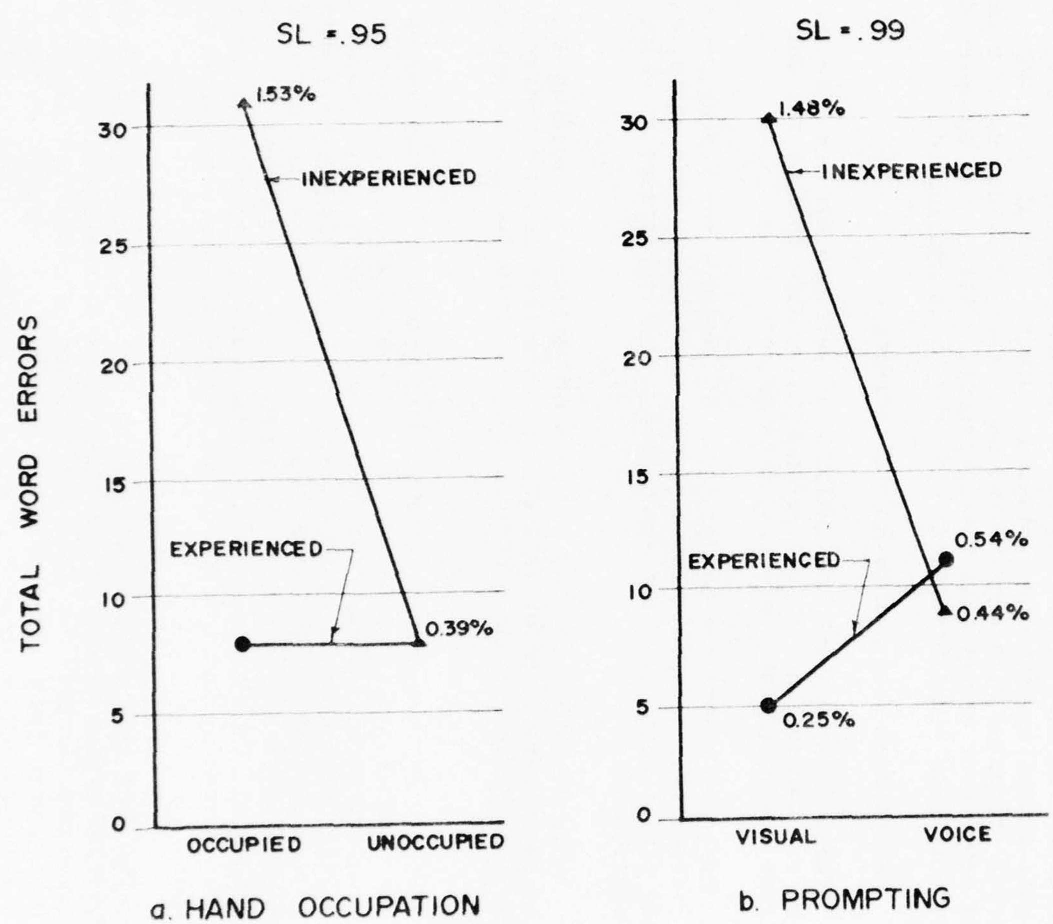


Fig. 21 Total Word Errors - Two Interactions With Experience

Figure 22 a) is a plot of reading and interpretation word errors versus hand occupation. Hand occupation produced four times as many errors as no hand occupation. Figure 20 b) shows the interaction between experience and hand occupation. For experienced subjects, the errors decreased slightly with hand occupation. For inexperienced subjects, the errors increased very significantly with hand occupation.

Figure 23 a) and b) are plots of interactions between experience and prompting and between hand occupation and prompting respectively for reading and interpretation word errors. Voice prompting produced many fewer errors than visual prompting with inexperienced subjects, and slightly more errors than visual prompting with experienced subjects. Likewise, voice prompting produced many fewer errors than visual prompting with hands occupied and slightly more than visual prompting with the hands unoccupied.

5. Analysis of Word Errors Before Correction

Table 22 summarizes the analysis of variance for the total word errors before correction. It can be seen that entry mode is the only really significant single factor affecting this measure.

Table 23 summarizes the analysis of variance for keying, recognition, and correction system (word) errors before correction. Once again, entry mode is highly significant and there are no other individual factors which are highly significant.

Table 24 summarizes the analysis of variance for reading and interpretation (word) errors before correction. For this measure, entry mode is not significant, and the one individual factor which is significant is prompting.

Figure 24 is a graph of word errors before correction versus entry mode. The graph shows the total errors and the breakdown into keying, recognition and correction system errors and into reading and interpretation errors. The number of total errors before correction is slightly more than four times as great for voice as for either of the other two entry modes. The number of keying, recognition and correction system errors for voice input is about ten times as great as for either of the other two entry modes. The number of reading and interpretation errors differs very little between entry modes.

The fact that voice input had ten times as many keying, recognition and correction system errors as the other two modes deserves some comment. Neither keyboard nor Graf Pen made any recognition errors nor did either device have a very large correction system error rate due to misrecognition of the correction system commands. Almost all keying errors except numerical keying errors were furthermore detected by the syntax checks of the data entry system, and were rejected. Voice entry, on the other hand, was relatively prone to recognition errors because the talkers were relatively inexperienced, training was abbreviated, and the data entry task was a stressful once.

In Figure 24, the three percentage error values which are given for each entry mode are based upon the total number of words to be entered by all

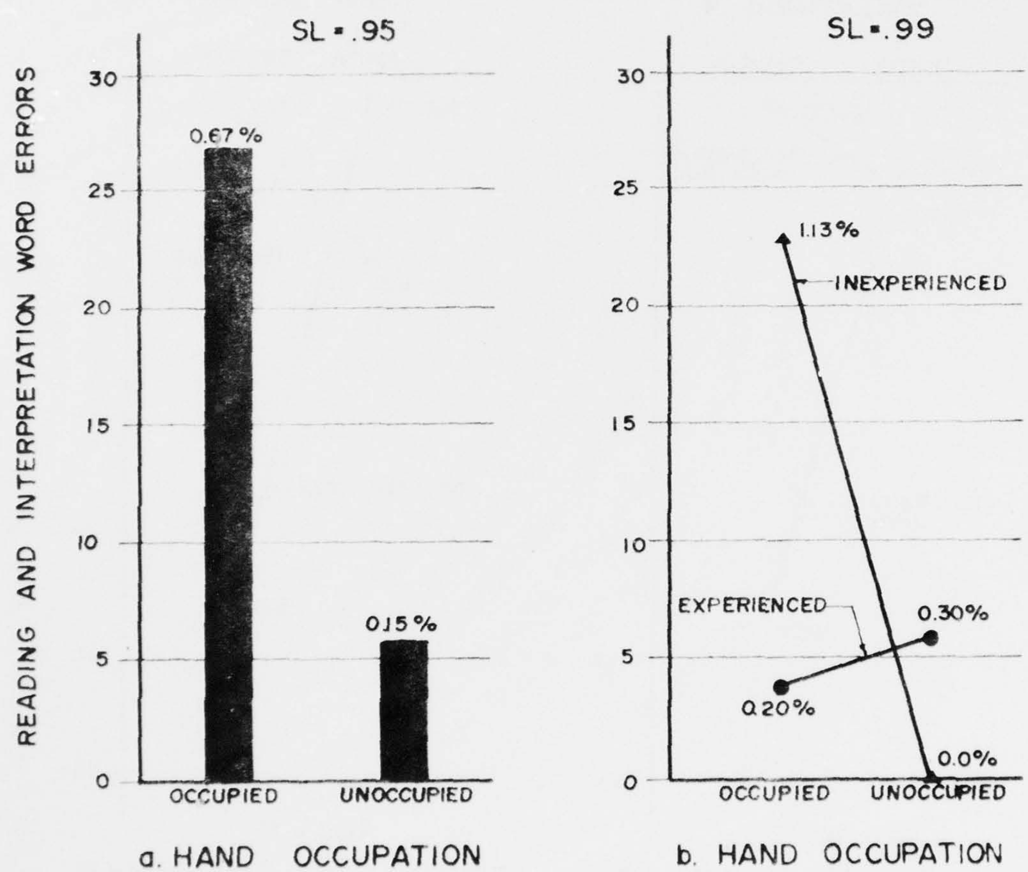


Fig. 22 Reading and Interpretation Word Errors Versus Hand Occupation and Experience

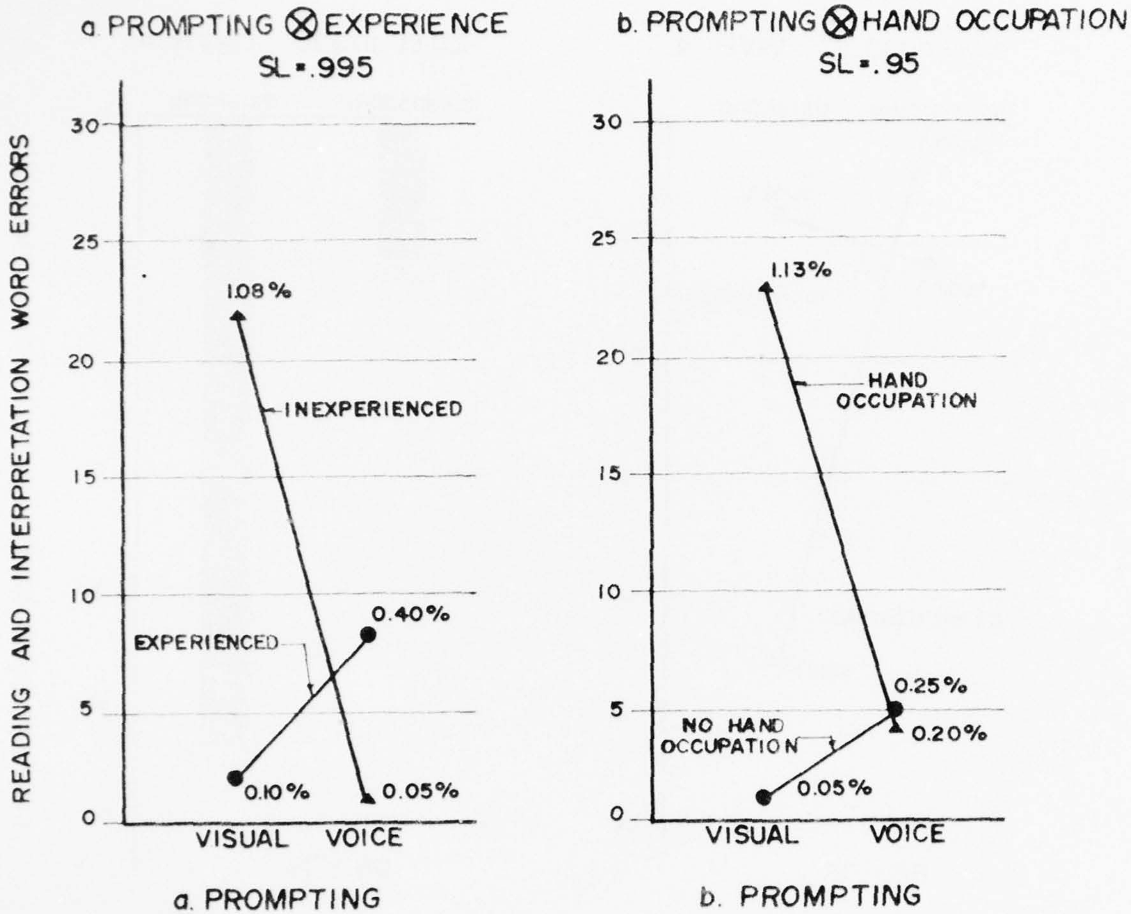


Fig. 23 Reading and Interpretation Word Errors - Two Interactions with Prompting Mode

TABLE 22

ANALYSIS OF VARIANCE OF
TOTAL WORD ERRORS BEFORE CORRECTION

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERIENCE (E)	1	0.056	0.056	-	-
2.	HAND OCCUPATION (H)	1	16.056	16.056	-	-
3.	ExH	1	0.500	0.500	-	-
4.	PROMPTING (P)	1	29.389	29.389	4.00	.90
5.	ExP	1	4.500	4.500	-	-
6.	HxP	1	1.389	1.389	-	-
7.	ENTRY MODE (M)	2	295.028	147.514	20.10	.9999
8.	ExM	2	4.694	2.347	-	-
9.	HxM	2	20.361	10.181	-	-
10.	PxM	2	10.028	5.014	-	-
11.	TRAINING (T)	2	23.028	11.514	-	-
12.	ExT	2	2.694	1.347	-	-
13.	HxT	2	73.861	36.931	5.03	.95
14.	PxT	2	10.194	5.097	-	-
15.	MxT	4	52.556	13.139	-	-
16.	ExHxM	2	50.583	25.292	3.45	.95
	ALL INTERACTIONS BETWEEN 3, 4 AND 5 FACTORS = ERROR	45	330.276	7.339		
	TOTAL	71	874.608			

GRAND MEAN = 2.639

TABLE 23

ANALYSIS OF VARIANCE OF
KEYING, RECOGNITION AND CORRECTION SYSTEM ERRORS BEFORE CORRECTION

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERIENCE (E)	1	1.681	1.681	-	-
2.	HAND OCCUPATION (H)	1	7.347	7.347	-	-
3.	ExH	1	15.125	15.125	4.34	.95
4.	PROMPTING (P)	1	5.014	5.014	-	-
5.	ExP	1	0.347	0.347	-	-
6.	HxP	1	0.125	0.125	-	-
7.	ENTRY MODE (M)	2	286.694	143.347	41.09	.99999
8.	ExM	2	4.528	2.264	-	-
9.	HxM	2	31.361	15.681	4.49	.95
10.	PxM	2	6.028	3.014	-	-
11.	TRAINING (T)	2	18.111	9.056	2.59	.90
12.	ExT	2	0.778	0.389	-	-
13.	HxT	2	17.444	8.722	2.50	.90
14.	PxT	2	10.111	5.056	-	-
15.	MxT	4	31.639	7.910	2.27	.90
	ALL INTERACTIONS BETWEEN 3, 4 AND 5 FACTORS = ERROR	45	156.983	3.489		
	TOTAL	71	593.317			

GRAND MEAN = 1.847

TABLE 24

ANALYSIS OF VARIANCE OF
READING AND INTERPRETATION ERRORS BEFORE CORRECTION

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERIENCE (E)	1	0.681	0.681	-	-
2.	HAND OCCUPATION (H)	1	2.347	2.347	-	-
3.	ExH	1	11.681	11.681	6.49	.95
4.	PROMPTING (P)	1	11.681	11.681	6.49	.95
5.	ExP	1	8.681	8.681	4.82	.95
6.	HxP	1	1.681	1.681	-	-
7.	ENTRY MODE (M)	2	1.694	0.847	-	-
8.	ExM	2	2.528	1.264	-	-
9.	HxM	2	5.861	2.931	-	-
10.	PxM	2	1.361	0.681	-	-
11.	TRAINING (T)	2	1.028	0.514	-	-
12.	ExT	2	2.528	1.264	-	-
13.	HxT	2	21.028	10.514	5.84	.99
14.	PxT	2	0.528	0.264	-	-
15.	MxT	4	10.639	2.660	-	-
16.	ExHxP	1	7.347	7.347	4.08	.95
	ALL INTERACTIONS BETWEEN 3, 4 AND 5 FACTORS = ERROR	45	81.042	1.80		
	TOTAL	71	164.986			

GRAND MEAN = 0.764

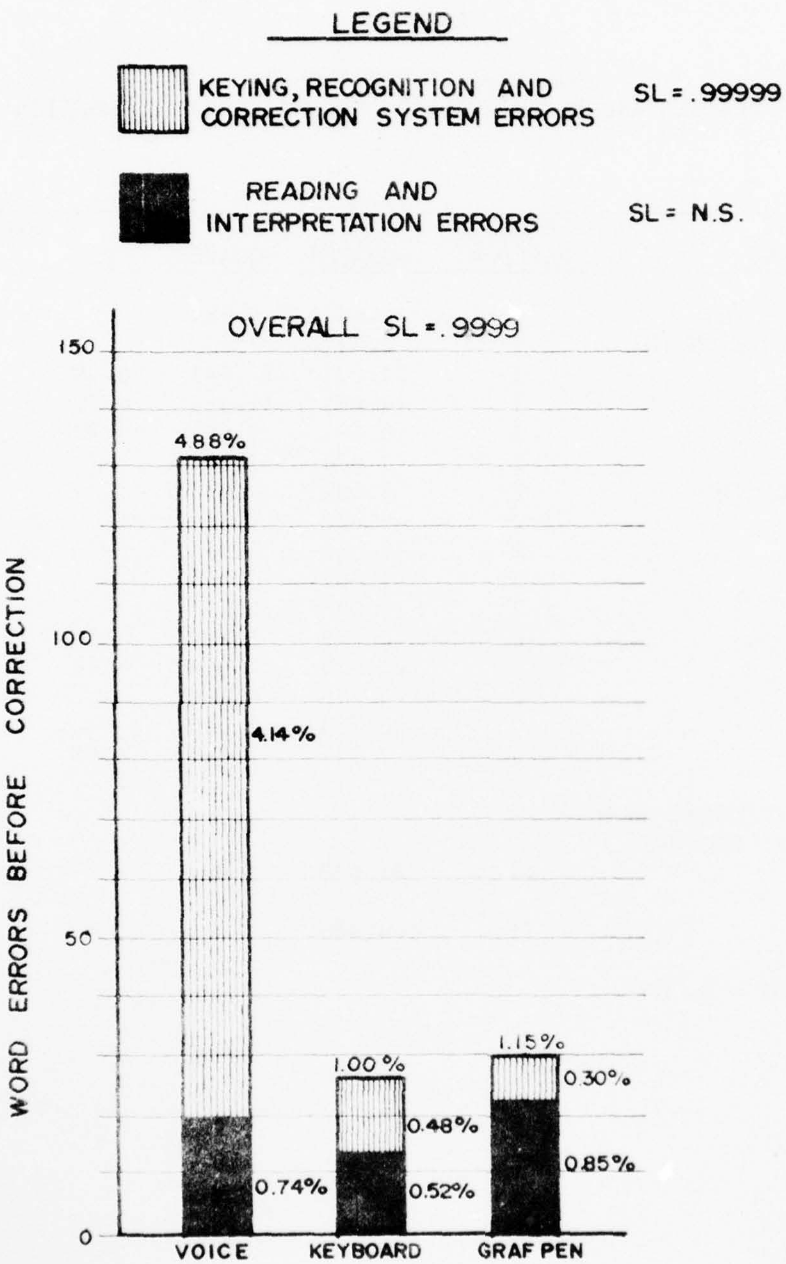


Fig. 24 Word Errors Before Correction Versus Entry Mode

subjects using that mode. These are not error rates per utterance as were measured in the High Speed Data Entry Test, and that partially accounts for the higher numerical values. Since each corrected word requires two additional utterances when a backspace correction is used and possibly an entire string of utterances when an erase correction is used, the actual error rates per utterance are probably on the order of 10 to 15% lower than these figures would indicate.

Figure 25 is a plot of reading and interpretation errors before correction versus prompting. The addition of voice prompting to visual prompting can be seen to reduce the incidence of this type of error to about one-third.

6. Analysis of Correction System Errors

Table 25 summarizes the analysis of variance for correction system errors. The errors counted here were always corrected before final verification. There may have been a few errors after correction which could have been attributable to correction system problems, but they were not broken down in the error counts and consequently are not included in this data. From Table 25, it can be seen that entry mode and an interaction between hand occupation and entry mode both have statistical significance with respect to corrected correction-system word errors.

Figure 26a and 26b are plots of correction system errors versus entry mode and versus an interaction between entry mode and hand occupation. There were 22 such errors with voice input, five with keyboard and only one with Graf Pen. Eighteen of the 22 voice input errors occurred with hand occupation.

7. Analysis of Rejects

Table 26 summarizes the analysis of variance for total system rejects in the high complexity data entry tests. Entry mode was barely significant at the 0.90 level. Hand occupation was significant at the 0.99 level.

Figure 27 is a plot of the number of rejects versus entry mode. Voice had about 50 rejects, keyboard had 30, and Graf Pen had 20. Figure 27b illustrates the relationship between hand occupation and the reject rate. Hand occupation resulted in about two and one-half times as many rejects as no hand occupation. The fact that the interaction between hand occupation and entry mode is not significant indicates that this is generally true for all three entry modes.

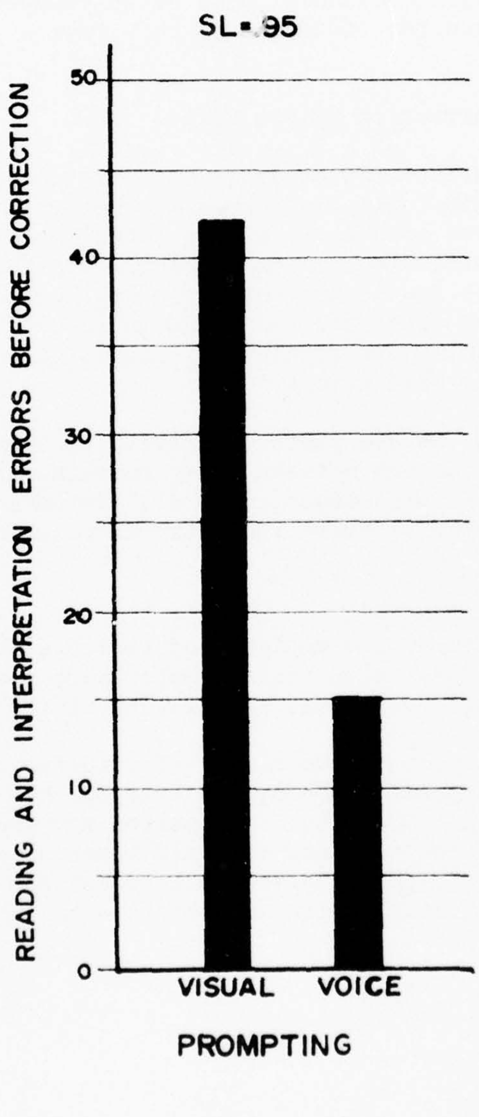


Fig. 25 Reading and Interpretation Errors Before Correction Versus Prompting

TABLE 25

ANALYSIS OF VARIANCE OF
CORRECTED CORRECTION-SYSTEM WORD ERRORS

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERIENCE (E)	1	0.500	0.500	-	-
2.	HAND OCCUPATION (H)	1	2.000	2.000	3.85	.90
3.	ExH	1	1.389	1.389	-	-
4.	PROMPTING (P)	1	0.056	0.056	-	-
5.	ExP	1	0.222	0.222	-	-
6.	HxP	1	0.056	0.056	-	-
7.	ENTRY MODE (M)	2	10.361	5.181	9.96	.999
8.	ExM	2	1.083	0.542	-	-
9.	HxM	2	6.250	3.125	6.01	.99
10.	PxM	2	0.194	0.097	-	-
11.	TRIAL (T)	2	1.861	0.931	-	-
12.	ExT	2	0.083	0.042	-	-
13.	HxT	2	0.583	0.292	-	-
14.	PxT	2	0.194	0.097	-	-
15.	MxT	4	2.889	0.722	-	-
	ALL INTERACTIONS BETWEEN 3, 4 AND 5 FACTORS = ERROR	45	23.4	0.52		
	TOTAL	71				

GRAND MEAN = 0.389

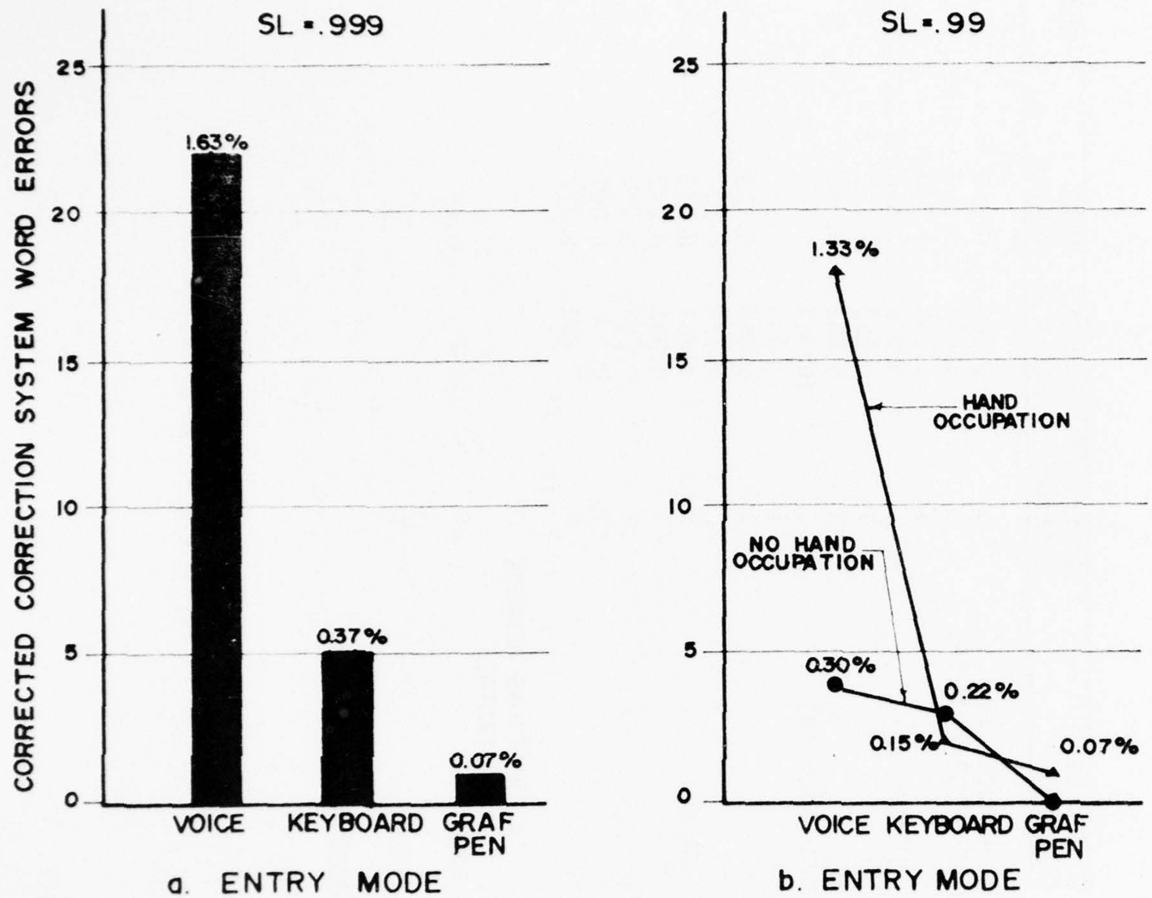


Fig. 26 Corrected Correction System Word Errors

TABLE 26

ANALYSIS OF VARIANCE OF
REJECTS

	SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F	SIGNIFICANCE LEVEL
1.	EXPERIENCE (E)	1	0.222	0.222	-	-
2.	HAND OCCUPATION (H)	1	26.889	26.889	8.38	.99
3.	ExH	1	0.222	0.222	-	-
4.	PROMPTING (P)	1	0.889	0.889	-	-
5.	ExP	1	5.556	5.556	-	-
6.	HxP	1	0.0	0.0	-	-
7.	ENTRY MODE (M)	2	15.528	7.764	2.42	.90
8.	ExM	2	14.694	7.347	-	-
9.	HxM	2	1.861	0.931	-	-
10.	PxM	2	15.194	7.597	-	-
11.	TRIAL (T)	2	11.194	5.597	-	-
12.	ExT	2	3.694	1.847	-	-
13.	HxT	2	1.194	0.597	-	-
14.	PxT	2	3.028	1.514	-	-
15.	MxT	4	14.639	3.660	-	-
	ALL INTERACTIONS BETWEEN 3, 4 AND 5 FACTORS = ERROR	45	144.23	3.21		
	TOTAL	71	259.111			

GRAND MEAN = 1.389

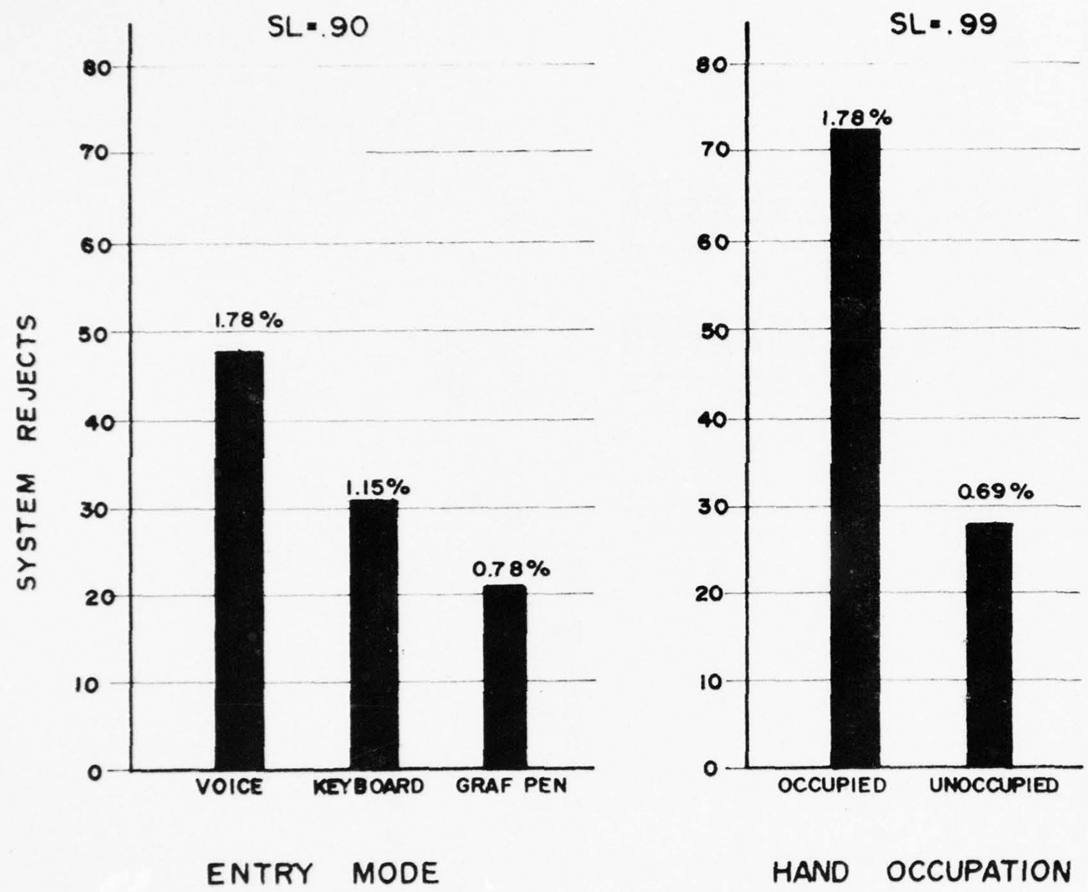


Fig. 27 Rejects Versus Entry Mode and Hand Occupation

Section IV

DISCUSSION OF RESULTS AND CONCLUSIONS

A. Discussion of HSDE Test Results

The High Speed Data Entry tests were a measure of data entry performance in a simple data copying problem. The subjects were skilled technical and office employees, most of whom were familiar with office machines such as typewriters, but who were not highly trained for these particular tests.

Three different performance measures were analyzed; average time per correct character, error rate (after correction), and error rate before correction. Error rate after correction is the operational error rate if it is assumed that verification and checking are allowed before the data is actually entered. Error rate before correction is indicative of the basic error rate of the entry device and the problem setting and would be the error rate in a system which did not allow for verification.

1. Entry Speed Comparisons

With respect to average time per correct character, a number of interesting statistically significant results were obtained.

Keyboard was clearly the fastest entry device, requiring an average of 29% less time per correct character than voice, and 22% less time than Graf Pen. It is important to note that most of the 16 keyboard subjects were familiar with the layout of the keys and two were expert typists. In the High Complexity Data Entry test, we have found that for subjects who are not familiar with the layout of the keys, keyboard tends to be a very slow entry device. For this kind of test and for subjects with some experience, however, keyboard is a fast and accurate data entry device.

A Graf Pen working in the menu mode is not as fast per character as a keyboard being used by a subject with some typing skill, since the Graf Pen forces the operator to work, at best, like a one-fingered typist. The Graf Pen would, however, regain the advantage if it were being used to enter entire words with one stroke from a well organized menu as compared to typing entire words or multicharacter abbreviations of words on the keyboard.

Voice was faster per character than the average times per correct character would indicate. A factor that slows down voice entry is the requirement to correct errors. It is possible to obtain less than 1% error rates with the VIP-100 voice recognizer when subjects who are fully trained are not striving for maximum entry speed, and when the system itself is trained with ten repetitions per word. In this experiment, however, the subjects were minimally trained, maximum entry speed was the objective, and only five repetitions per word were used for training the voice input system. Consequently, the error rates went up to 2 or 3%. Since the subjects were also striving for accuracy, they had to stop and make corrections whenever the recognizer made an error. The time to make these corrections significantly increased the average entry time. Some subjects did manage to find an optimum entry speed (somewhat below maximum speed) which resulted in a lower error rate and hence, a higher overall entry rate, but variability introduced into the voice input response time by errors in the priority interrupt structure made it difficult for most subjects to perform this kind of optimization.

One of the greatest surprises of the test is that the addition of voice response feedback to visual feedback had no significant effect on the speed of data entry, and in fact, only affected high speed data entry by producing higher error rates under some circumstances. This result was surprising since conceptually, voice response feedback would seem to provide the advantage of freeing the eyes from the verification process. In practice, however, the voice response unit which we used had two problems (neither of which was necessarily inherent to it as a particular voice response unit). First it was too slow. If a feedback device is to be useful for high speed data entry, it must be fast. There is a reason to doubt that any voice response system could be fast enough for feedback in high speed data entry, but since we definitely did not test a fast VRU, our data cannot be used to support this conclusion. The second problem with the VRU which we used was that it had a small suboptimum vocabulary. The words which were used for entry and verification of alphabetic data were dictated by the VRU and were neither easy to remember nor particularly natural for entry of alphabetic data.

A second surprising result was that hand occupation had no overall significant effect on speed of data entry. In interaction with entry mode, hand occupation did provide some discrimination since it slowed down Graf Pen more than keyboard and was accompanied by an increase in entry speed for voice input.

Hand occupation had very little effect because it was too simple a task. This was demonstrated by the fact that the requirement to push buttons affected entry speed differently for different types of data. For alphanumeric input or input of 10-character-long strings (for which pushing the buttons consumed only a small fraction of the total time) pushing the buttons increased throughput slightly (possibly by improving rhythm or adding dicipline to the entry task). For the case of short numeric input strings (for which the time to push the buttons was a more significant fraction of the total time), the pushbuttons reduced the throughput as expected.

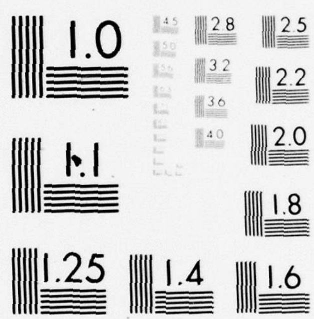
The experimental factor with the highest statistical significance was the data entry character set. Entry time was 25% less for numeric data than it was for alphanumeric data. One reason for this is that a smaller vocabulary reduces the time to find keys or Graf Pen menu locations, or to recall voice entry code words. A second reason is that the smaller vocabulary reduces the error rate and associated correction time for machine errors and human errors.

A final factor which affected entry speed was the length of the character strings. Overall 10-character strings required 14% less time per character than 3-character strings. The fact that long strings require substantially less overhead time for verification than short strings accounts for this difference, but this effect is partially cancelled by the fact that long

AD-A045 939 THRESHOLD TECHNOLOGY INC DELRAN N J F/G 17/2
AUTOMATIC DATA ENTRY ANALYSIS.(U)
SEP 77 J R WELCH F30602-76-C-0178
UNCLASSIFIED RADC-TR-77-306 NL

2 OF 2
ADA045939

END
DATE
FILMED
12-77
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

strings cannot be memorized as a whole and must be entered in three or more separate parts.

These effects are demonstrated more clearly by considering the (significant) interaction between string length and input mode. For voice input, the verification process involved speaking an additional word which constituted a substantial overhead for 3-character strings. For keyboard and Graf Pen input, verification introduced somewhat less overhead for 3-character strings. On the other hand, entry of 10-character strings can be accomplished by voice without breaking the string into parts and can be accomplished by keyboard without losing touch with the keys, but cannot be accomplished on Graf Pen without requiring at least partial reorientation to the menu locations. As a result, voice input time was substantially less for 10-character strings than for 3-character strings; keyboard time was slightly less, and Graf Pen time was actually slightly greater for 10-character strings.

2. Operational Error Rate (Errors After Correction)

The only two parameters which had significant overall effect on error rate were the data string length and the data entry alphabet (character set). The effect of string length was trivial, however, since it was significant only for string errors. For these errors, the error rate was roughly proportional to the string length, as would be expected.

The effect of character set was more interesting. The error rate was approximately twice as high for alphanumeric data as it was for numeric data for both character errors and string errors. In addition, there was a significant interaction between the alphabet and entry mode. For alphanumeric data, the operational error rate was lower for voice entry than for either keyboard or Graf Pen. For numeric data, however, the error rate was lowest for keyboard entry, slightly higher with Graf Pen and substantially higher for voice.

Let us now consider why alphanumeric data had a higher overall error rate and favored voice input, while numeric data had a lower overall error rate and favored keyboard and Graf Pen input.

Operational errors consisted of two principal components; display reading errors, and entry device (recognition or keying) errors. For alphanumeric data, the display reading errors were substantial, and consisted of confusions between I and 1, and S and 5, and occasional data deletions. These errors occurred frequently with both keyboard and Graf Pen entry since with these devices the subjects' eyes were occupied with finding keys and menu positions. On the other hand, voice entry resulted in a lower reading error rate, probably because voice entry does not require use of the eyes. With all numeric data, confusion between similar characters was not likely and the predominant errors were recognition and keying errors. Since the recognition error rate for voice is higher than the keying error rates for keyboard or Graf Pen, voice entry had the highest error rate with this type of data. Consequently, it appears that since voice data entry frees the eyes as well as the hands, it has the effect of reducing the number of uncorrected reading errors. If the data entry alphabet is complex or the system is one that would give rise to reading errors for other reasons, voice entry has an accuracy advantage. If reading errors are not a problem, voice entry loses its accuracy advantage by virtue of its higher recognition error rate.

Feedback mode had no significant overall effect on operational accuracy but it did have a significant interaction with alphabet. In particular, the addition of voice response feedback substantially degraded operational accuracy for alphanumeric data but produced a slight improvement in accuracy with numeric data. It is possible that the limited vocabulary of the voice response unit may have caused this effect. For numeric data, the voice response unit fed back the names of the characters directly as displayed. For alphabetic data, however, the VRU fed back a set of almost arbitrary words that were related to the characters only by their first letters. The results indicate that this kind of loosely related feedback can do more harm than good.

3. Errors Before Correction

Errors before correction may have little relationship to the output error pattern of a data entry system, but they have substantial bearing upon the internal design and the efficiency of the system. For example, entry mode was not significant with respect to errors after correction since all three entry devices had about the same average error rate. It was significant with respect to errors before correction, however, because voice input had about twice the error rate of the other two devices. In addition, roughly two errors were being corrected with voice input for each error which remained after correction.

Hence, one explanation is provided for why voice input was not as fast as the other two entry devices. It was strikingly clear to the author when observing the HSDE tests that a major factor affecting speed of entry with voice input was the requirement to correct recognition errors. Furthermore, the frequency of correction was only part of the problem since often in the process of making corrections with the voice input system, further errors were generated either by misrecognition of the correction commands or by misrecognition of the new entry data. It seemed that errors begot errors, possibly because of the disturbing effect that they had upon the subjects in the relatively high pressure environment of the tests. Conversely, the effects of minor improvements in basic recognition performance would tend to be magnified into even greater improvements in overall system performance.

One other interesting result with respect to errors before correction was that the addition of voice response feedback had virtually no effect on the error rates for Graf Pen or voice entry, but increased the error rate by a factor of five for keyboard entry. We believe that the low speed of the voice response unit may have contributed to this higher error rate. The voice response unit was so slow relative to the keyboard that almost all subjects typed ahead of the feedback and tried to ignore it. It is possible that hearing the names of previously entered characters spoken while trying to enter a new character may have produced confusion that resulted in reading, memory, and keying errors.

B. Discussion of HCDE Test Results

The high complexity data entry tests were a measure of data entry performance in a complicated data entry setting in which the subject's ability to interpret an English language statement and convert it to a series of data entry fields had as much effect upon data entry performance as did the raw speed of the data entry system.

The subjects were skilled technical and office employees and were divided into two classes depending upon their experience levels with the particular data entry device. The subjects were not highly trained for these particular data entry tests, however, so that the experiment is indicative of performance rates which would be achieved by casual users of a data entry system. With more training, the relationships between some factors would possibly change and the overall performance levels definitely would improve.

Six different performance measures were analyzed; entry time per word, field errors, word errors, word errors before correction, corrected correction system errors, and rejects. In addition, the field and word error measures were broken down into reading and interpretation errors, and keying, recognition and correction system errors.

1. Entry Speed Comparisons

Overall, voice was the fastest entry mode in these tests. Graf Pen required an insignificant average of 5% more time per word than voice, and keyboard required a highly significant average of 29% more time per word. There was, in addition, a very significant interaction between entry mode and subject experience. For experienced subjects, the three devices were nearly identical in speed. For inexperienced subjects, the entry time increased only 5% for voice and 14% for Graf Pen but jumped 56% for keyboard. The significantly higher entry time for inexperienced keyboard subjects is an indication of how difficult it is to search for characters on a completely unfamiliar teletypewriter keyboard. The time difference was also magnified by the fact that all of the non-numeric entry words required two characters on the keyboard but only one entry with voice or Graf Pen.

It is interesting to consider why the Graf Pen didn't require more time per word than it did, since it had a 43 word menu that was completely new to both experienced and inexperienced subjects. The Graf Pen times were not particularly high, primarily because the menu was organized specifically for the particular data entry problem being tested. At each stage of the data entry process it was only necessary to find the data row indicated by the prompting messages and then to scan that row for the proper entry. If the Graf Pen had been set up as a light pen with only the applicable segments of the menu being displayed at each stage of the entry hierarchy, its entry time almost certainly would have been reduced further, since the requirement to interpret prompts and search for data rows would have been eliminated.

In a like manner, it is probable that both voice and keyboard entry would have been faster if instead of prompting with names of data fields, prompting had been done by displaying lists of the acceptable input responses at each entry stage. The advantages of this approach would, of course, decrease as the length of the lists increased.

Hand occupation affected entry speed in these tests in a generally predictable way. The 3.5 second button pushing requirement increased entry times significantly on an overall basis. As in the case of the high speed data entry tests, hand occupation had the greatest effect with Graf Pen (a 30% time increase), less effect with keyboard (a 20% increase) and the least effect with voice (a 9% increase). With no hand occupation, Graf Pen was faster than voice in these experiments, but by a statistically insignificant amount.

In these tests, voice response was used for prompting but was not used explicitly for feedback. Once again, voice response surprisingly failed to make a statistically significant overall impact on entry speed. It did, however, achieve significance in interaction with entry mode. The addition of voice prompting increased entry time for voice input by about 18%, had virtually no effect on keyboard, and decreased entry time by about 9% for the Graf Pen.

Voice response prompting slowed down voice data entry because most subjects waited for the VRU to stop talking before they would start talking. The subjects never seemed to be compelled to try to achieve higher throughput by getting ahead of the VRU as they did in the high speed data entry tests, possibly because entering data per se was only a part of the total problem in these tests. Here again, a much faster VRU would have provided a performance advantage.

On the other hand, with keyboard and Graf Pen input, no one hesitated to enter data while the voice response unit was talking. It was also clear that the Graf Pen subjects were actually using the voice response unit to relieve them of the requirement for reading prompts and verifying non-numeric entries. Keyboard subjects didn't generally find the VRU as useful because visual prompting was more conveniently located for them than for the Graf Pen subjects and because experienced keyboard subjects did not have their eyes fully occupied by the task of finding keys.

2. Operational Errors (Errors After Correction)

Errors after correction were analyzed in terms of field errors and word errors. Generally, these two ways of looking at the errors produced similar results, except that since each number in a numeric field was counted as a separate word there were more word errors. In particular, out of a total of 1080 test sentences there were 35 field errors and 55 word errors.

We have further subdivided the counts of field and word errors into two classes, which we will simply call recognition errors and reading errors. Recognition errors actually consisted of all keying, recognition, and correction system errors. Reading errors consisted of all reading and problem interpretation errors. Most of the errors which were counted as word errors, but were not counted as field errors were classified as reading errors since these were the kinds of errors which occurred primarily with numeric fields. For simplicity, the remaining discussion in the section will apply exclusively to the counts of word errors. The relationships for field errors do not differ substantially.

There were no significant differences in total word errors between the three entry devices. There were, however, about four times as many recognition errors with voice input as for the average of the other two entry modes and this result was statistically significant. On the other hand, keyboard and Graf Pen produced an average of slightly more than twice as many reading errors, but this result was not statistically significant. Overall, reading errors outnumbered recognition errors by a three to two ratio.

These results are consistent with the results from the high speed data entry tests, in that the three devices produced about the same number of operational errors except that voice entry produced mostly recognition errors and the other two devices produced mostly reading errors.

Both experience and hand occupation had significant average effects on word errors. The error rate was about 2.5 times greater for inexperienced subjects than for experienced subjects, and was also about 2.5 times greater with hand occupation than without hand occupation. These differences were, moreover, completely related through an interaction. Hand occupation and lack of experience by themselves produced low error rates, but the combination of hand occupation and inexperience resulted in nearly a fourfold increase in error rate. This relationship was true for total word errors and for reading errors but not for recognition errors.

A similar interaction existed for experience and prompting. The error rates were relatively low for all combinations of these two variables except for the case of inexperienced subjects using visual prompting. This combination resulted in about a threefold increase in total errors and reading errors.

Prompting and hand occupation also interacted strongly with respect to reading errors. The combination of visual prompting and hand occupation resulted in about a five to one increase in error rate as compared to the other three combinations of these two variables. This result was not true for total word errors or for recognition errors.

These interactions may or may not be meaningful. They are all significant at the 0.95 level or higher, but since the total number of errors is so small, the performance of one or two subjects could easily bias the overall results. It does seem clear, however, that inexperience, hand occupation, and lack of voice response prompting had a tendency to increase reading and interpretation errors. The nature of the interactions, furthermore, indicates that there may have been a threshold effect. Any of the adverse conditions by themselves did not result in increased error rates, but all combinations of two adverse conditions gave rise to substantial increases in reading error rates.

3. Word Errors Before Correction

Word errors before correction provide an indication of the basic error performance of the entry systems. These errors were also broken down into two classes; keying, correction system, and recognition errors, and reading and problem interpretation errors.

The primary factor which was significant with respect to total word errors before correction was entry mode. Voice entry had about four times as many errors before correction as either of the other entry devices. Furthermore, voice entry required correction of more than five errors for every error that remained after correction.

The primary factor which was significant with respect to keying, recognition, and correction system errors was, once again, entry mode. Voice entry produced nearly ten times as many of these errors as either of the other devices. This difference occured because neither keyboard nor Graf Pen made any recognition errors. Almost all keying errors, except numerical keying errors, were furthermore detected by the syntax checks of the data entry system and were rejected. Hence, the keying and recognition error rate was very low for these devices. Voice entry, on the other hand, was relatively prone to recognition errors since in these tests, the talkers were either completely inexperienced or only moderately experienced, training was abbreviated (only five samples per word with little retraining), and the data entry task was relatively stressful.

The only individual factor which was significant with respect to reading and interpretation errors before correction was prompting. The addition of voice prompting to visual prompting reduced the incidence of this type of error to about one-third. Evidently, the additional prompting helped the subjects to determine which data fields they should extract at each entry point. This is one of the few situations in which voice response has provided the exact advantage which would be expected of it.

4. Corrected Correction System Errors

Entry mode, hand occupation and an interaction between entry mode and hand occupation were the only factors which were significant with respect to corrected correction system errors. Voice had about four times as many of these errors as keyboard and about twenty times as many as Graf Pen, which only had one such error.

Correction system errors were easily observed by the author while conducting the tests. Human errors and machine errors relative to the use of the backspace and erase commands were particularly disconcerting with voice input because they reduced the entry rate, and often confused the subject so much that he would make further recognition or correction system errors. The correction system was very clear for Graf Pen entry (a back arrow for backspace and the word erase for erasing the entire entry). The correction system was not as clear for keyboard (rubout for backspace and shift-rubout for erasure), so that its slightly higher error rate might have been anticipated.

For voice input, the words backspace and erase did not mean the same things to all subjects and were often erroneously used, particularly by experienced subjects who were accustomed to other words for these functions. In addition, since both words were acceptable to the syntax at almost all times, there were numerous false recognitions of those words. False recognition of the erase command near the end of an otherwise correct message was particularly disconcerting.

The correction system error rate with voice input was about four times higher with hand occupation than it was without hand occupation. Since voice was the only entry mode which allowed the hand occupation requirement to be fulfilled while data was being entered, the higher incidence of correction system errors during hand occupation could be an indication of additional stress produced by simultaneous voice data entry and hand occupation.

5. Rejects

The most significant experimental factor affecting the reject rate was hand occupation. Hand occupation resulted in about two and one-half times as many rejects as no hand occupation. The fact that the interaction between hand occupation and entry mode is not significant indicates that this is generally true for all three entry modes. Hence, it appears that hand occupation produced a form of stress which resulted in a greater incidence of illegal or garbled entries.

The differences in reject rate for the three entry modes were barely significant. Voice had about fifty rejects, keyboard had thirty, and Graf Pen had twenty. Voice and keyboard may have produced more rejects than Graf Pen because they had multiple reject modes. Voice would reject on erroneous entries or misrecognition of the correct entry. Keyboard would reject on erroneous entries or if either keystroke of a two-letter entry was in error. Graf Pen would reject only on an illegal entry.

C. Capsule Summary of Results

1. HSDE Tests

a. Entry Speed

- Keyboard was the fastest device overall (most subjects had at least some keyboard experience); Graf Pen was slower; voice was the slowest.

- Voice response feedback added to visual feedback had no significant effect on entry speed.

- The instantaneous two-handed pushbutton requirement had no significant overall effect on entry speed, but did slow Graf Pen entry slightly, keyboard somewhat less, and actually was accompanied by a slight increase in voice entry speed.

- The greatest slow-down effect from hand occupation was with entry of 3-character numeric strings.

- Alphanumeric entry required 25% more time than numeric entry and was the most significant experimental factor affecting speed.
- Overall, 10-character strings were entered faster than 3-character strings. The difference was the greatest for voice input, and less for keyboard. For Graf Pen, entry of 3-character strings was slightly faster than entry of 10-character strings.

b. Operational Error Rate

- Long strings had higher string error rates than short strings.
- The alphanumeric data set had about twice the error rate of the numeric data set.
- For alphanumeric data, voice input had a lower error rate than keyboard or Graf Pen. For numeric data, voice had a higher error rate than either keyboard or Graf Pen.
- The addition of voice response feedback degraded accuracy for alphanumeric data, but had little effect on numeric data.

c. Errors Before Correction

- Voice input had about twice the before correction error rate of either keyboard or Graf Pen.
- Voice response feedback had virtually no effect on the before correction error rate of voice input or Graf Pen but increased the error rate five-fold for keyboard.

2. HCDE Tests

a. Entry Speed

- Voice and Graf Pen were fastest in this test. Keyboard required 29% more time per word than voice. The higher time for keyboard was all attributable to inexperienced subjects. For them, input time was 56% greater than for experienced subjects.
- Hand occupation slowed Graf Pen most, keyboard less and voice least, and had a significant overall effect on entry speed.
- Voice response prompting had no significant overall effect but it slowed input by voice significantly, slightly increased entry speed for Graf Pen and had no effect on keyboard.

b. Operational Error Rate

- There were no significant differences between the three devices in total operational errors, but voice had mostly recognition errors while keyboard and Graf Pen had mostly reading errors.
- The combination of inexperience and hand occupation greatly increased the operational error rate, mostly due to reading errors.

- The combination of inexperience and lack of voice prompting greatly increased the operational error rate- mostly due to reading errors.

- The combination of hand occupation and lack of voice response prompting greatly increased the operational error rate for reading errors only.

c. Errors Before Correction

- Voice input had about four times as many errors before correction as either of the other two entry devices, primarily because it had about ten times as many keying, recognition and correction system errors.

- The before correction reading and interpretation error rates of the three devices were not significantly different.

- The addition of voice response prompting reduced the reading and interpretation error rate by a factor of three.

d. Correction System Errors

- Voice input had four times as many of these errors as keyboard and twenty times as many as Graf Pen.

- Most of the correction system errors with voice input occurred with hand occupation.

e. Rejects

- Hand occupation increased the reject rate by a factor of two and one-half.

D. Conclusions

1. Voice Data Entry

Voice data entry has demonstrated some advantages in these tests which go beyond the obvious advantages which it has when the hands are fully occupied. In a simple data copying scenario which was prone to reading errors, it provided a lower error rate than keyboard or Graf Pen. In a complex data entry scenario requiring substantial mental and visual effort, it provided a higher throughput than keyboard, particularly with inexperienced subjects. In both cases, the advantages accruing from voice input were almost certainly related to its ability to free the subject's eyes from the task of finding keys or menu locations.

Voice entry also had some problems. In a simple task involving copying of alphabetic and/or numeric characters, the isolated word recognition system could not compete with keyboard or Graf Pen in terms of entry speed. Voice entry speed was limited by the requirement to pause between words, by additional small delays due to a software error, by its relatively higher error rate, and by the relatively great difficulty associated with correction of errors. For alphanumeric data, the lower speed was compensated to some extent by the greater entry accuracy which voice provided, but for numeric-only data, Graf Pen and keyboard were superior to voice in both entry speed and accuracy. For voice to provide an advantage for simple numeric data entry, either the hands would have to be very busy or recognition would have to be provided for rapidly spoken continuous digits.

In the high complexity scenario, voice entry had a relatively high but not excessive error rate before correction. Voice entry proceeded smoothly in comparison with the other two entry modes, was on the average faster, and had an insignificantly higher error rate after correction. A lower basic error rate, however, would almost certainly have made voice entry even faster and would have provided an even greater demonstration of the advantages of "eyes free" data entry in a complex scenario.

The conclusions which derive from these results include the obvious recommendation for reducing the error rate and response time of voice data entry systems. In addition, however, these results suggest that an undue emphasis on voice as a hand-freeing data entry mode may be obscuring its possibly more important advantages as an eye-freeing, mind-freeing, data entry device which is particularly suitable for use by individuals without keyboard training.

If, for example, it were possible to combine some simple manual control functions with the voice input process and thereby increase both recognition accuracy and response time, the improved data entry performance would, for many applications, substantially outweigh the disadvantages imposed by the requirement to use the hands. In particular, these experiments have shown that correction system errors contribute substantially to the entry time and higher error rates for voice input. Hence, the simple addition of a set of well marked correction keys to work in parallel with the spoken correction commands could produce a significant improvement in error rate and entry speed for inexperienced users. Other functions which could be put under manual control include the verification command and possibly even the signals which indicate the boundaries of the words to be recognized. This latter possibility could conceivably provide some of the same speed advantages as would be provided by a continuous speech recognition system, but at a much lower cost.

2. Keyboard Entry

These experiments have indicated that keyboard provides rapid, accurate data entry of simple strings of characters when used by subjects with some keyboard experience. For entry of small vocabularies of words, it loses some of its speed advantage as compared to voice or numeric oriented entry because of the requirement for striking several keys per word. It also suffers a remarkable reduction in speed when used by totally inexperienced subjects. In this respect, it differs from voice or Graf Pen entry, since for those devices totally inexperienced operators were not much slower than operators with hours of experience.

Keyboard accuracy was adversely affected in these tests by the addition of character-by-character voice response feedback, because the feedback was too slow to keep up with the entry device and the subjects were confused by the feedback of previously entered character names.

The correction system commands used for keyboard entry in these tests were slightly ambiguous, so that some correction system errors were made. Keyboard data entry systems which provide for instantaneous correction or deletion of data should have clearly marked easily accessible keys for those purposes.

3. Menu Data Entry

Menu data entry in these experiments was not quite as fast as keyboard for entry of simple strings of characters, probably because menu data entry was, at best, like "one-fingered" typing. For data entry of primarily words in a more complex scenario, however, the menu was faster than keyboard for inexperienced users. Menu entry was accomplished in this case by single strokes on a menu tailored specifically to the entry scenario.

In both data entry tests, hand occupation caused a greater speed reduction for menu data entry than for keyboard or voice input, probably because part of the entry system had to be transported back and forth between the menu and the hand occupation pushbuttons in the menu mode but not in the other modes.

In the high complexity data entry test, voice response prompting slightly increased entry speed for menu entry. It did not increase entry speed for the other two devices.

Correction system errors were almost non-existent for menu entry apparently because the correction system menu markings were self-explanatory, graphically related to the functions they performed, and easy to find.

Menu data entry could be highly recommended for situations involving entry of medium sized vocabularies of words in either a simple or a complex scenario with no hand occupation and with availability of voice response prompting.

A further refinement which would almost certainly improve the speed and accuracy of menu data entry would be to display the menus on a CRT, and to configure the selection system to work in a light pen mode. This would allow for variable menus and larger vocabularies and would probably obviate any requirement for voice response prompting.

4. Voice Response

Voice response feedback of individual characters in the simple data entry tests had no significant effect on entry speed, and no significant overall effects on entry accuracy. This feedback did, however, have a substantial negative effect on accuracy of alphanumeric data entry. It also produced a large increase in the number of errors before correction for keyboard entry.

Most of these effects are related to the relatively slow talking speed and limited vocabulary of the particular VRU that was used in these tests. The addition of voice response feedback had no effect on entry speed because the VRU was slower than any of the entry devices, but was fully buffered so that the subjects could continue to enter data even though the VRU had not finished talking. With keyboard input, the subjects went much faster than the VRU and left it far behind. We believe that the high error rate before correction with keyboard input and voice response feedback may have resulted from hearing feedback of past data while entering current data. The high error rate for alphanumeric data relative to numeric data probably resulted from the choice of feedback words. For numeric data, the words were the numbers themselves. For alphabetic data, the words were related to the alphabetic data only through correspondence of initial letters.

In the high complexity data entry tests, voice response prompting had no significant overall effect on entry speed but did affect speed differently for different entry modes. It reduced voice input speed substantially because the subjects stopped entering data when the VRU was talking. The VRU did not affect entry speed by keyboard and increased entry speed slightly for Graf Pen.

Voice response was decisively beneficial in these tests only in its effect on reading and interpretation errors in the high complexity data entry tasks. It had a strong tendency to reduce such errors both before and after correction. This affect was the one clear demonstration that voice response, like voice input, can free the eyes from at least one of the burdens of the data entry task and can thereby improve data entry performance. The fact that voice response only showed this advantage in the prompting mode agrees with Hammerton's[1] conclusion that instructions should be heard and data seen.

5. Hand Occupation

These experiments have provided some surprising results with respect to hand occupation. Hand occupation does tend to favor voice input, but the advantages are only significant when the hand occupation time is a substantial fraction of the total entry time. A momentary hand occupation task alternating with several seconds of data entry is virtually insignificant in discriminating entry modes.

In addition to slowing data entry by keyboard and Graf Pen, hand occupation tended to increase certain kinds of errors for all entry devices. In the high complexity data entry tests, addition of hand occupation seemed to be a stress factor which increased reading and interpretation errors and the reject rate for all three entry devices. In addition, for voice input, hand occupation greatly increased the number of correction system errors.

---

[1] Hammerton, M., "The Use of Same or Different Sensory Modalities in Information and Instructions", Royal Naval Personnel Research Committee Report, December 1974, AD-A026857.

Appendix A

AUTOMATIC DATA ENTRY ANALYSIS EXPERIMENTAL DATA

The experimental data for the high speed data entry tests is tabulated in Tables A-1 through A-4.

The experimental data for the high complexity data entry tests is tabulated in Tables A-5 through A-16.

TABLE A-1

HSDE RESULTS - AVERAGE TIME PER CORRECT CHARACTER

TEST NO.	MODE	ALPHABET	LENGTH	HANDS OCCUPIED	VOICE RESPONSE	T1	T2	T3
1	V	N	3	Y	N	2.21	1.79	1.90
2					Y	4.27	2.39	1.81
3				N	N	1.86	1.28	1.10
4					Y	2.59	1.93	1.98
5			10	Y	N	1.21	.84	.78
6					Y	1.48	1.34	1.11
7				N	N	1.64	2.12	1.14
8					Y	1.73	1.18	1.60
9		A/N	3	Y	N	2.52	2.14	1.81
10					Y	2.48	2.17	1.71
11				N	N	2.01	2.67	2.14
12					Y	2.73	2.32	1.98
13			10	Y	N	1.89	1.39	1.33
14					Y	1.56	1.29	1.79
15				N	N	2.11	2.03	1.90
16					Y	2.80	1.50	1.77
17	K	N	3	Y	N	1.23	1.11	1.09
18					Y	1.92	2.09	1.06
19				N	N	1.24	.98	.88
20					Y	.99	.89	.90
21			10	Y	N	1.30	1.20	1.07
22					Y	1.16	.86	.81
23				N	N	1.07	.99	1.00
24					Y	.96	.72	.66
25		A/N	3	Y	N	1.73	1.28	1.13
26					Y	3.41	1.72	1.31
27				N	N	1.68	1.48	1.40
28					Y	2.51	1.97	1.72
29			10	Y	N	1.28	1.17	1.19
30					Y	2.15	1.24	1.25
31				N	N	2.09	1.71	1.46
32					Y	1.84	1.31	1.08
33	G	N	3	Y	N	2.29	1.74	1.51
34					Y	1.85	1.40	1.44
35				N	N	1.02	.95	.97
36					Y	1.48	1.21	1.15
37			10	Y	N	1.45	1.32	1.59
38					Y	2.20	1.24	1.23
39				N	N	2.16	1.22	1.38
40					Y	1.98	1.28	1.21
41		A/N	3	Y	N	2.07	2.18	1.51
42					Y	1.92	1.75	1.88
43				N	N	2.86	2.12	1.93
44					Y	1.67	1.49	1.31
45			10	Y	N	2.32	2.03	1.98
46					Y	1.90	2.23	1.85
47				N	N	2.16	1.82	1.63
48					Y	1.91	1.85	1.62

A-2

TABLE A-2
HSDE RESULTS - PERCENT CORRECT CHARACTERS

TEST NO.	MODE	ALPHABET	LENGTH	HANDS OCCUPIED	VOICE RESPONSE	T1	T2	T3
1	V	N	3	Y	N	93.33	100	100
2					Y	100	97.33	100
3				N	N	100	100	100
4					Y	100	100	100
5			10	Y	N	100	100	98
6					Y	100	100	100
7				N	N	99	97	99
8					Y	94	98	98
9		A/N	3	Y	N	100	100	100
10					Y	97.33	100	98.66
11				N	N	97.33	98.66	100
12					Y	100	98.66	97.33
13			10	Y	N	99	100	99
14					Y	99	99	99
15				N	N	100	99	98
16					Y	100	99	100
17	K	N	3	Y	N	100	100	100
18					Y	96	100	100
19				N	N	100	100	100
20					Y	100	100	100
21			10	Y	N	100	100	100
22					Y	100	100	100
23				N	N	100	100	100
24					Y	100	100	100
25		A/N	3	Y	N	100	100	100
26					Y	93.33	97.33	96
27				N	N	100	98.66	98.66
28					Y	97.33	97.33	98.66
29			10	Y	N	97	100	100
30					Y	99	97	100
31				N	N	100	99	100
32					Y	96	100	100
33	G	N	3	Y	N	100	100	100
34					Y	100	100	100
35				N	N	100	100	100
36					Y	100	100	100
37			10	Y	N	100	100	100
38					Y	100	100	100
39				N	N	90	99	99
40					Y	100	100	100
41		A/N	3	Y	N	98.66	97.33	100
42					Y	97.33	100	98.66
43				N	N	93.33	98.66	100
44					Y	97.33	92	96
45			10	Y	N	98	100	99
46					Y	96	99	99
47				N	N	100	100	100
48					Y	99	100	100

TABLE A-3
HSDE RESULTS - PERCENT CORRECT CHARACTER STRINGS

TEST NO.	MODE	ALPHABET	LENGTH	HANDS OCCUPIED	VOICE RESPONSE	T1	T2	T3
1	V	N	3	Y	N	88	100	100
2					Y	100	92	100
3				N	N	100	100	100
4					Y	100	100	100
5			10	Y	N	100	100	90
6					Y	100	100	100
7				N	N	90	70	90
8					Y	60	80	90
9		A/N	3	Y	N	100	100	100
10					Y	96	100	96
11				N	N	92	96	100
12					Y	100	96	92
13			10	Y	N	90	100	90
14					Y	90	90	90
15				N	N	100	90	80
16					Y	100	90	100
17	K	N	3	Y	N	100	100	100
18					Y	92	100	100
19				N	N	100	100	100
20					Y	100	100	100
21			10	Y	N	100	100	100
22					Y	100	100	100
23				N	N	100	100	100
24					Y	100	100	100
25		A/N	3	Y	N	100	100	100
26					Y	80	92	88
27				N	N	100	96	96
28					Y	92	92	96
29			10	Y	N	70	100	100
30					Y	90	70	100
31				N	N	100	90	100
32					Y	60	100	100
33	G	N	3	Y	N	100	100	100
34					Y	100	100	100
35				N	N	100	100	100
36					Y	100	100	100
37			10	Y	N	100	100	100
38					Y	100	100	100
39				N	N	70	90	90
40					Y	100	100	100
41		A/N	3	Y	N	96	92	100
42					Y	92	100	96
43				N	N	80	96	100
44					Y	92	84	88
45			10	Y	N	80	100	90
46					Y	70	90	90
47				N	N	100	100	100
48					Y	90	100	100

TABLE A-4

HSDE RESULTS - PERCENT WRONG CHARACTERS BEFORE CORRECTION (PER UTTERANCE)

TEST NO.	MODE	ALPHABET	LENGTH	HANDS OCCUPIED	VOICE RESPONSE	T1	T2	T3
1	V	N	3	Y	N	4.5	2.59	.86
2					Y	.93	2.83	0
3				N	N	6.83	0	0
4					Y	6.50	3.63	2.72
5			10	Y	N	1.48	1.75	1.81
6					Y	1.63	1.58	.87
7				N	N	2.45	6.85	0.9
8					Y	6.72	2.67	4.2
9		A/N	3	Y	N	0	0	0
10					Y	4.72	3.51	2.77
11				N	N	.97	5.30	3.70
12					Y	4.88	4.59	1.92
13			10	Y	N	6.15	2.5	4.06
14					Y	0.83	.9	2.45
15				N	N	2.45	1.69	6.55
16					Y	1.40	.9	1.73
17	K	N	3	Y	N	0	0	0
18					Y	2.94	1.75	0
19				N	N	0	0	0
20					Y	.98	0	.98
21			10	Y	N	.89	0	0
22					Y	3.31	2.58	.89
23				N	N	.89	0	0
24					Y	.85	0	0
25		A/N	3	Y	N	0	0	0
26					Y	5.45	3.77	3.0
27				N	N	0	1.0	1.0
28					Y	2.80	1.98	.99
29			10	Y	N	2.72	0	.89
30					Y	5.11	2.72	0
31				N	N	.84	1.69	0
32					Y	5.80	.89	0
33	G	N	3	Y	N	0	0	0
34					Y	0	0	0
35				N	N	0	0	0
36					Y	0	0	0
37			10	Y	N	0	0	2.38
38					Y	1.67	0	0
39				N	N	10.56	.84	2.10
40					Y	3.79	3.76	2.56
41		A/N	3	Y	N	.90	4.91	1.88
42					Y	1.98	.96	.91
43				N	N	5.0	1.0	0
44					Y	2.0	6.6	2.97
45			10	Y	N	1.81	0	.9
46					Y	3.63	2.52	.9
47				N	N	.87	0	0
48					Y	.9	.89	0

TABLE A-5

HCDE RESULTS - ENTRY TIME PER WORD

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPER-IENCE	T1	T2	T3
1	V	VI	N	E	2.47	1.35	1.39
2				I	2.17	1.84	1.82
3			Y	E	3.13	2.35	1.72
4				I	2.49	2.56	1.80
5		VO	N	E	2.64	2.20	2.07
6				I	3.66	2.20	2.16
7			Y	E	3.30	2.07	1.83
8				I	2.85	2.38	2.00
9	K	VI	N	E	2.21	1.72	1.48
10				I	4.15	3.19	3.14
11			Y	E	3.56	2.68	2.26
12				I	4.26	3.68	3.27
13		VO	N	E	2.38	1.99	1.58
14				I	3.89	3.18	3.12
15			Y	E	3.32	2.41	2.05
16				I	4.40	3.48	3.19
17	G	VI	N	E	2.34	1.91	1.58
18				I	3.04	2.56	2.43
19			Y	E	3.01	2.61	2.20
20				I	3.28	2.55	2.34
21		VO	N	E	2.17	1.67	1.47
22				I	2.21	1.79	1.48
23			Y	E	3.03	2.60	2.04
24				I	3.56	2.92	2.35

N.C.

TABLE A-6

HCDE RESULTS - TOTAL FIELD ERRORS

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPER-IENCE	T1	T2	T3
1	V	VI	N	E	0	0	2
2	V	VI	N	I	3	0	1
3	V	VI	Y	E	1	0	0
4	V	VI	Y	I	1	2	1
5	V	VO	N	E	0	0	0
6	V	VO	N	I	0	0	2
7	V	VO	Y	E	0	1	0
8	V	VO	Y	I	2	0	2
9	K	VI	N	E	0	0	0
10	K	VI	N	I	0	1	0
11	K	VI	Y	E	1	0	0
12	K	VI	Y	I	1	2	0
13	K	VO	N	E	1	0	0
14	K	VO	N	I	0	0	0
15	K	VO	Y	E	0	2	0
16	K	VO	Y	I	0	0	0
17	G	VI	N	E	0	0	0
18	G	VI	N	I	0	0	0
19	G	VI	Y	E	0	1	0
20	G	VI	Y	I	1	0	1
21	G	VO	N	E	0	0	2
22	G	VO	N	I	0	0	0
23	G	VO	Y	E	0	0	1
24	G	VO	Y	I	1	2	0

H-7

TABLE A-7

HCDE RESULTS - KEYING, RECOGNITION AND CORRECTION SYSTEM FIELD ERRORS

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPER-IENCE	T1	T2	T3
1	V	VI	N	E	0	0	1
2				I	3	0	1
3			Y	E	1	0	0
4				I	1	1	0
5		VO	N	E	0	0	0
6				I	0	0	2
7			Y	E	0	1	0
8				I	2	0	2
9	K	VI	N	E	0	0	0
10				I	0	1	0
11			Y	E	1	0	0
12				I	0	0	0
13		VO	N	E	1	0	0
14				I	0	0	0
15			Y	E	0	0	0
16				I	0	0	0
17	G	VI	N	E	0	0	0
18				I	0	0	0
19			Y	E	0	0	0
20				I	0	0	0
21		VO	N	E	0	0	0
22				I	0	0	0
23			Y	E	0	0	1
24				I	1	1	0

TABLE A-8

HCDE RESULTS - READING AND INTERRPRETATION FIELD ERRORS

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPER-IENCE	T1	T2	T3
1	V	VI	N	E	0	0	1
2				I	0	0	0
3			Y	E	0	0	0
4				I	0	1	1
5		VO	N	E	0	0	0
6				I	0	0	0
7			Y	E	0	0	0
8				I	0	0	0
9	K	VI	N	E	0	0	0
10				I	0	0	0
11			Y	E	0	0	0
12				I	1	2	0
13		VO	N	E	0	0	0
14				I	0	0	0
15			Y	E	0	2	0
16				I	0	0	0
17	G	VI	N	E	0	0	0
18				I	0	0	0
19			Y	E	0	1	0
20				I	1	0	1
21		VO	N	E	0	0	2
22				I	0	0	0
23			Y	E	0	0	0
24				I	0	1	0

TABLE A-9

HCDE RESULTS - TOTAL WORD ERRORS

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPER-IENCE	T1	T2	T3
1	V	VI	N	E	0	0	2
2				I	3	0	1
3			Y	E	1	0	0
4				I	2	4	1
5		VO	N	E	0	0	0
6				I	0	0	2
7			Y	E	0	1	0
8				I	2	0	2
9	K	VI	N	E	0	0	0
10				I	0	2	0
11			Y	E	1	0	0
12				I	5	5	0
13		VO	N	E	1	0	0
14				I	0	0	0
15			Y	E	0	3	0
16				I	0	0	0
17	G	VI	N	E	0	0	0
18				I	0	0	0
19			Y	E	0	1	0
20				I	5	0	2
21		VO	N	E	0	0	5
22				I	0	0	0
23			Y	E	0	0	1
24				I	1	2	0

TABLE A-10

HCDE RESULTS - KEYING, RECOGNITION AND CORRECTION SYSTEM WORD ERRORS

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPERIENCE	T1	T2	T3
1	V	VI	N	E	0	0	1
2				I	3	0	1
3			Y	E	1	0	0
4				I	1	1	0
5		VO	N	E	0	0	0
6				I	0	0	2
7			Y	E	0	1	0
8				I	2	0	2
9	K	VI	N	E	0	0	0
10				I	0	2	0
11			Y	E	1	0	0
12				I	0	0	0
13		VO	N	E	1	0	0
14				I	0	0	0
15			Y	E	0	0	0
16				I	0	0	0
17	G	VI	N	E	0	0	0
18				I	0	0	0
19			Y	E	0	0	0
20				I	0	0	0
21		VO	N	E	0	0	0
22				I	0	0	0
23			Y	E	0	0	1
24				I	1	1	0

TABLE A-11

HCDE RESULTS - READING AND INTERPRETATION WORD ERRORS

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPER-IENCE	T1	T2	T3
1	V	VI	N	E	0	0	1
2				I	0	0	0
3			Y	E	0	0	0
4				I	1	3	1
5		VO	N	E	0	0	0
6				I	0	0	0
7			Y	E	0	0	0
8				I	0	0	0
9	K	VI	N	E	0	0	0
10				I	0	0	0
11			Y	E	0	0	0
12				I	5	5	0
13		VO	N	E	0	0	0
14				I	0	0	0
15			Y	E	0	3	0
16				I	0	0	0
17	G	VI	N	E	0	0	0
18				I	0	0	0
19			Y	E	0	1	0
20				I	5	0	2
21		VO	N	E	0	0	5
22				I	0	0	0
23			Y	E	0	0	0
24				I	0	1	0

TABLE A-12

HCDE RESULTS - TOTAL WORD ERRORS BEFORE CORRECTION

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPER-IENCE	T1	T2	T3
1	V	VI	N	E	2	2	12
2				I	4	1	10
3			Y	E	17	6	6
4				I	5	11	4
5		VO	N	E	3	1	1
6				I	12	0	4
7			Y	E	11	6	3
8				I	6	3	2
9	K	VI	N	E	0	0	1
10				I	2	3	1
11			Y	E	1	0	0
12				I	5	5	0
13		VO	N	E	1	1	0
14				I	0	0	0
15			Y	E	2	4	0
16				I	1	0	0
17	G	VI	N	E	2	1	4
18				I	0	0	3
19			Y	E	0	1	0
20				I	5	1	3
21		VO	N	E	1	0	6
22				I	0	0	0
23			Y	E	0	0	1
24				I	1	2	0

TABLE A-13

HCDE RESULTS - READING AND INTERPRETATION WORD ERRORS BEFORE CORRECTION

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPER-IENCE	T1	T2	T3
1	V	VI	N	E	0	0	4
2				I	0	0	5
3			Y	E	2	0	0
4				I	1	3	1
5		VO	N	E	0	0	0
6				I	2	0	0
7			Y	E	2	0	0
8				I	0	0	0
9	K	VI	N	E	0	0	0
10				I	0	0	0
11			Y	E	0	0	0
12				I	5	5	0
13		VO	N	E	0	0	0
14				I	0	0	0
15			Y	E	0	3	0
16				I	1	0	0
17	G	VI	N	E	2	1	3
18				I	0	0	0
19			Y	E	0	1	0
20				I	5	1	3
21		VO	N	E	0	0	6
22				I	0	0	0
23			Y	E	0	0	0
24				I	0	1	0

TABLE A-14

HCDE RESULTS - KEYING, RECOGNITION AND CORRECTION SYSTEM
WORD ERRORS BEFORE CORRECTION

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPERIENCE	T1	T2	T3
1	V	VI	N	E	2	2	8
2				I	4	1	5
3			Y	E	15	6	6
4				I	4	8	3
5		VO	N	E	3	1	1
6				I	10	0	4
7			Y	E	9	6	3
8				I	6	3	2
9	K	VI	N	E	0	0	1
10				I	2	3	1
11			Y	E	1	0	0
12				I	0	0	0
13		VO	N	E	1	1	0
14				I	0	0	0
15			Y	E	2	1	0
16				I	0	0	0
17	G	VI	N	E	0	0	1
18				I	0	0	3
19			Y	E	0	0	0
20				I	0	0	0
21		VO	N	E	1	0	0
22				I	0	0	0
23			Y	E	0	0	1
24				I	1	1	0

TABLE A-15

HCDE RESULTS - CORRECTED CORRECTION SYSTEM WORD ERRORS

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPER-IENCE	T1	T2	T3
1	V	VI	N	E	0	1	0
2				I	0	0	0
3			Y	E	4	1	2
4				I	1	3	0
5		VO	N	E	1	0	0
6				I	2	0	0
7			Y	E	1	3	1
8				I	0	2	0
9	K	VI	N	E	0	0	0
10				I	2	0	0
11			Y	E	0	0	0
12				I	0	0	0
13		VO	N	E	0	1	0
14				I	0	0	0
15			Y	E	2	0	0
16				I	0	0	0
17	G	VI	N	E	0	0	0
18				I	0	0	1
19			Y	E	0	0	0
20				I	0	0	0
21		VO	N	E	0	0	0
22				I	0	0	0
23			Y	E	0	0	0
24				I	0	0	0

TABLE A-16

HCDE RESULTS - REJECTS

TEST NO.	MODE	PROMPTING MODE	HAND OCCUPATION	EXPERIENCE	T1	T2	T3
1	V	VI	N	E	2	3	1
2				I	1	2	1
3			Y	E	8	6	5
4				I	1	2	0
5		VO	N	E	0	0	1
6				I	3	0	0
7			Y	E	1	2	2
8				I	2	4	1
9	K	VI	N	E	0	0	0
10				I	3	1	0
11			Y	E	2	0	2
12				I	6	2	1
13		VO	N	E	1	2	0
14				I	1	2	0
15			Y	E	4	0	0
16				I	2	0	2
17	G	VI	N	E	0	0	0
18				I	2	0	0
19			Y	E	0	2	0
20				I	0	0	1
21		VO	N	E	1	0	1
22				I	0	0	0
23			Y	E	0	1	1
24				I	1	10	1

Appendix B

DESCRIPTION OF THE VIP-100 VOICE RECOGNITION SYSTEM

The VIP-100 automatic speech recognition system is a product of Threshold Technology Inc., 1829 Underwood Boulevard, Delran, New Jersey 08075. The VIP-100 recognizes words spoken in isolation and can be automatically adapted for different speakers and/or vocabularies. The system can be trained on-line and provides, as an output, a digital code which can be used to enter data into a computer, retreive stored information, or control machine operations.

The basic VIP-100 system consists of four units; a preprocessor, a minicomputer, an output display and a Teletype. The preprocessor accepts the speech input from the microphone and converts it to logic signals which are then processed by the (Nova 1200) minicomputer. The computer compares the input signal with stored references to determine which, if any, of the vocabulary words were spoken. If a correlation is found between the input speech and one of the vocabulary words, an appropriate message will be sent to the output display; a reject indicator will be lighted if no correlation is found. The Teletype is used for control, and for input and output functions.

Before an operator uses the VIP-100 in the recognition mode, the system is first optimized for the particular vocablary and for the operator's manner of speaking by the use of a training routine. The operator speaks several utterances of each word during training. After training, the VIP-100 can recognize the chosen vocabulary words when they are spoken by the operator that trained the system. It is not necessary to retrain the system each time a different operator uses the system since the training data may be stored in computer memory or on punched paper tape. The appropriate tape with the stored data can be read into the system whenever the operator or vocabulary is changed The system may be retrained for a single word, multiple words, or the complete vocabulary at any time in order to accommodate vocabulary word substitutions or temporary changes in an operator's speech characteristics which may result from colds or other respiratory ailments.

In the recognition mode, response time to the spoken words is virtually instantaneous and recognition outputs can be printed using the Teletype or visually observed on a display. Forced decisions can be made or "no decision" threshold criteria can be established, thereby requiring the speaker to repeat his utterance before a word decision is made.

Specification for the VIP-100 are given in Table B-1. The system will operate to specifications in machine noise backgrounds as high as 85-90 dB. A variety of options are available depending upon the system application requirements.

Table B-1

VIP-100 SPECIFICATIONS

Vocabulary	- Up to 32 discrete words or short phrases
Training	- On-line or off-line. Training less than 10 seconds per word. Paper tape input/output of speakers' training data.
Operation	- Response time less than 0.1 seconds. Minimum spacing between words 0.1-0.2 seconds. Storage of multi-speaker training data.
Output	- Digital encoded output. Visual display of recognition results. Hard copy teletype printout.
Physical	- Basic Hardware: Power 115 VAC, single phase, 60 Hz, 500 watts, weight 120 pounds, size 18 x 20 x 26 inches. Standard ASR 33 Teletype. Standard visual display.
Options	- Vocabulary expandable to 100 words. Telephone interface. Output: Voice response, hard copy, special visual display, special purpose control, processing, statistical computations, custom output interfacing. Custom turnkey system design. Off-line loading of training data. Hardware rack mountable.

Appendix C

DESCRIPTION OF GRAF PEN SONIC DIGITIZER

The Graf Pen GP-3 Sonic Digitizer is a product of Science Accessories Corporation, 970 Kings Highway West, Southport, Connecticut 06490. The system employed in this experiment included the basic GP-3 control unit, a standard ball-point stylus, a pair of point sensor microphones, and an interface board for a Data General Nova minicomputer. The general specifications for the Graf Pen GP-3 are given in Table C-1.

Table C-1

GENERAL SPECIFICATIONS FOR GRAF PEN GP-3

Resolution
for English units...0.01 inch

Data Rate...........Variable up to 140 points per second for 14-inch sensors, decreases slightly with longer sensors.

Reproducibility......0.1% of full scale or ± least significant bit, whichever is greater.

Digital Outputs
Registers..........X and Y (up to 13-bit) binary or 4-digit BCD, with standard TTL buffers (also line drivers and open collector buffers available).

Output Ready.......Ground-going pulse
Pen Control........Ground when stylus is in contact with display surface

Controls (on front panel)
Power.............On/off
Rate...............Coordinate-pair rate selectable up to 140 points a second
Modes.............Point, line, run and remote
Left Hand/Right Hand.........Set by push-pull operation of "rate" knob.

Indicators (on front panel)
X and Y displays (optional).............Two groups of four or five digits
Power..................................On
Left Hand..............................."L"
Stylus outside sensor area..............."M"

Connectors
Front Panel........For stylus or cursor cable
Back panel........Output connector for X data and Y data, X and Y register overflows, output-ready (program interrupt), pen control (Z axis), external reset, and sensors

Tablet (standard)....Useful area 14" x 14" (other sizes available, clear or frosted acrylic or phenolic-surfaced hardboard.

Appendix D

DESCRIPTION OF STC MODEL 200 VOICE GENERATOR

The STC Model 200 Voice Generator is a product of Speech Technology Corporation, 631 Wilshire Boulevard, Santa Monica, California 90401. The capabilities of the Model 200 have been achieved by combining a high-quality, small inexpensive, solid-state voice synthesizer with highly compressed digital vocabularies which are programmed into read-only memories (ROMs) within the voice generator. This data compression allows 30 to 40 spoken words (30 seconds of continuous speech) to be stored in four standard 8K ROMs. The ROMs are interfaced with self-contained logic circuitry to select any message in the vocabulary - whether it is a word, a phrase, or a sentence - from short binary-coded input signals. Space is provided in the Model 200 for additional vocabulary storage - up to a total of 200 to 250 words or two minutes of speech.

The voice generator contains its own power supply and has two independent voice signal outputs. One output is at standard telephone-line level, and the other is at 0.5 watts into 8-ohm speaker load. A limited amount of DC power is also furnished for external TTL, MOS, and CMOS control circuits.

The specifications for the Model 200 voice generator are given in Table D-1.

Table D-1

SPECIFICATIONS FOR STC MODEL 200 VOICE GENERATOR

Mechanical -	Size: 6.2 x 2.0 x 10.5 inches, exclusive of mounting feet and panel projections - panel size is 6.2 x 2.0 inches Weight: 7.0 lbs. (15 lbs. shipping weight).
Environment-	0^{o} to 40^{o} C operating, -50^{o} to + 100^{o} C storage, 0 to 90% operating relative humidity, non-condensing.
Power -	115V $\pm$ 10%, 50 to 400 Hz, 15 W maximum
Signal Connections-	(DBM-25S connector Data signals TTL compatible)

The Model 200 Voice Generator is shipped with:

. Desk-top, 8-ohm loudspeaker
. Speaker connection cord
. Power cord
. DBM-25P signal connector
. Reference manual

MISSION
of
Rome Air Development Center

RADC plans and conducts research, exploratory and advanced development programs in command, control, and communications (C^3) activities, and in the C^3 areas of information sciences and intelligence. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.

AMERICAN REVOLUTION BICENTENNIAL
1776 · 1976

Printed by
United States Air Force
Hanscom AFB, Mass. 01731